PRACE NAUKOWE AKADEMII EKONOMICZNEJ WE WROCŁAWIU
Nr 1126 2006
TAKSONOMIA 13
Klasyfikacja i analiza danych – teoria i zastosowania

Dorota Rozmus
Akademia Ekonomiczna w Katowicach

BADANIE ZWIĄZKÓW MIĘDZY MIARAMI ZRÓŻNICOWANIA BŁĘDÓW POJEDYNCZYCH MODELI A JAKOŚCIĄ MODELU ZAGREGOWANEGO

1. Wstęp

Zastosowanie modeli zagregowanych w klasyfikacji i regresji cieszy się w ostatnich latach bardzo dużym zainteresowaniem. Niejednokrotnie modele te wykazują lepsze właściwości teoretyczne i empiryczne niż modele pojedyncze. Jakość modelu zagregowanego w silnym stopniu zależy od jakości pojedynczych modeli. Jednakże, jak pokazują badania empiryczne, nie jest to jedyny czynnik mający na nią wpływ. Okazuje się bowiem, że specyficzne relacje zachodzące między wartościami teoretycznymi pojedynczych modeli w silnym stopniu wpływają na błąd modelu zagregowanego. W literaturze nie ma zgody co do tego, jaki charakter powinny mieć związki między składowymi modelu zagregowanego. Jak piszą Ruta i Gabrys [2002b, s. 689]: „Zróżnicowanie, niezależność, niezgodność, a coraz częściej i ujemna korelacja, są terminami często stosowanymi, by określić pożądaną relację między pojedynczymi modelami; taką, która zapewniłaby wysoką jakość modelu zagregowanego". Relacja ta uwidacznia się w ten sposób, że błędne wartości teoretyczne dla różnych obserwacji nie będą powstawały na podstawie tych samych modeli.

Teoretycznie, grupa niezależnych modeli przynosi poprawę jakości w stosunku do jakości modeli pojedynczych. Dowodzą tego m.in. Hansen i Salamon [1990]. Zauważyli oni, że aby błąd modelu zagregowanego, budowanego na zasadzie głosowania majoryzacyjnego[1], był niższy niż błąd pojedynczych modeli, konieczne jest, by składowe modelu zagregowanego różniły się między sobą i były „poprawne". Przez model poprawny rozumieją taki model, którego błąd jest mniejszy niż w

[1] Klasyfikując nowy obiekt, jako ostateczną wybiera się tę klasę, która najczęściej wskazywana była przez pojedyncze modele.

przypadku losowego przydzielania do klas. Natomiast dwa modele są różne, jeśli rozpoznając nowe obiekty niepoprawnie klasyfikują różne obserwacje. Zakładając, że prawdopodobieństwo popełnienia błędu na podstawie każdego pojedynczego modelu jest takie samo, równe p, i że błędy modeli są niezależne, to prawdopodobieństwo błędu modelu zagregowanego dla pojedynczego obiektu będzie równe:

$$P(y_n \neq \hat{y}_m) = \sum_{k>\frac{K}{2}}^{K} \binom{K}{k} p^k (1-p)^{(K-k)}, \tag{1}$$

gdzie: K to liczba pojedynczych modeli,

y_n – rzeczywista wartość zmiennej objaśnianej,

$\hat{y}_m$ – wartość teoretyczna modelu zagregowanego.

Wiadomo, że zróżnicowanie jest warunkiem koniecznym dla wysokiej jakości modelu zagregowanego, jednak nie ma zgody co do tego, jak je definiować i jak mierzyć. Znane są liczne sposoby jego pomiaru, jednak niewiele wiadomo o relacjach między miarami zróżnicowania i jakością modelu zagregowanego oraz o wyższości jednego typu miar nad innymi, co powoduje, że trudno wskazać, która z nich jest najlepsza.

Głównym oczekiwaniem związanym z miarami zróżnicowania jest to, że będą one pomocne w konstrukcji jak najlepszego modelu zagregowanego, tj. we właściwym wyborze metod konstrukcji pojedynczych modeli, liczby składowych modelu globalnego, operatorów ich agregacji itp.

Istnieją miary dla par modeli (*pairwise*), które liczone są dla wszystkich par pojedynczych modeli, a następnie uśredniane, oraz miary globalne (*non-pairwise*), które stosują miarę entropii, wariancji albo też oparte są na rozkładzie liczby błędnie sklasyfikowanych obserwacji.

Poza podziałem na miary globalne i miary dla par modeli miary te można także podzielić na:

- ujemnie skorelowane z błędem modelu zagregowanego; są to te, które śledzą zróżnicowanie: im wyższą wartość przybiorą, tym większe jest zróżnicowanie;
- dodatnio skorelowane z błędem modelu zagregowanego; są to takie, które śledzą podobieństwo: im wyższą wartość przyjmą, tym mniejsze jest zróżnicowanie.

W dalszej części artykułu skupiono się jedynie na miarach globalnych. Wszystkie one wymagają uprzedniego przekodowania wyników klasyfikacji na wartość 0, gdy na podstawie modelu uzyskano poprawną klasyfikację, i na wartość 1 w przypadku przeciwnym.

2. Uogólnione miary zróżnicowania

Entropia (*entropy measure*)

Jest to miara zastosowana w pracy [Kuncheva, Whitaker 2001]. Pokazuje stopień niezgodności między wartościami teoretycznymi pojedynczych modeli:

$$EN = \frac{1}{N}\sum_{n=1}^{N}\frac{\min\{m(\mathbf{x}_n),\ K - m(\mathbf{x}_n)\}}{K - \left\lceil \frac{K}{2} \right\rceil}, \tag{2}$$

gdzie: $m(\mathbf{x}_n)$ to liczba modeli, na podstawie których dokonano błędnej klasyfikacji n-tej obserwacji:

$$m(\mathbf{x}_n) = \sum_{n=1}^{N} I\{y_n \neq \hat{y}_k\}, \tag{3}$$

N – liczba obserwacji,

$\lceil a \rceil$ – najmniejsza liczba całkowita, większa niż a.

Entropia osiąga wartość najwyższą (EN=1), gdy modele są maksymalnie zróżnicowane między sobą; najniższą natomiast (EN = 0), gdy wszystkie wartości teoretyczne pojedynczych modeli są identyczne.

Log-entropia (*log-entropy*)

Należy zaznaczyć, że miara dana wzorem (2) nie jest klasyczną entropią ze względu na to, iż pomija funkcję logarytmiczną. W klasycznej formie została przedstawiona przez P. Cunnighama i J. Carneya [2000]:

$$lEN = \frac{1}{N}\sum_{n=1}^{N}\sum_{i=1}^{I} -\frac{N_i^n}{K}\log_I\left(\frac{N_i^n}{K}\right), \tag{4}$$

gdzie N_i^n to liczba modeli składowych, które zaklasyfikowały n-tą obserwacje do i-tej klasy.

Miara trudności (*measure of difficulty*)

Miara ta początkowo została zaproponowana przez L.K. Hansena i P. Salamona [1990], a następnie rozwinięta w pracy [Kuncheva, Whitaker 2003]. Budowana jest na podstawie tzw. dyskretnego rozkładu błędów (*discrete error distribution*) $V = [p_V(0), p_V(1),..., p_V(K)]$, zdefiniowanego przez D. Rutę i B. Gabrysa [2002a]. Miara ta określa wariancję elementów $p_V(k)$ odzwierciadlających znormalizowaną liczbę obserwacji, dla których zaobserwowano dokładnie k błędów. Jest definiowana jako:

$$DI = \frac{1}{K}\sum_{k=0}^{K}\left(p_V(k) - \overline{p}_V\right)^2. \tag{5}$$

Wariancja Kohavi i Wolperta (*Kohavi-Wolpert variance*)

Kolejna oprócz DI miara oparta na wariancji została przedstawiona przez P. Kohavi i D.H. Wolperta [1996]. Mierzy przeciętną wariancję binominalnego rozkładu wartości teoretycznych pojedynczych modeli:

$$KW = \frac{1}{NK^2}\sum_{n=1}^{N}\left[m(\mathbf{x}_n)\left(K - m(\mathbf{x}_n)\right)\right]. \tag{6}$$

Niejednoznaczność (*ambiguity*)

Oparta na wariancji jest także, zdefiniowana początkowo na potrzeby regresji przy wykorzystaniu kwadratowej funkcji straty przez A. Krogha i J. Vedelsby'ego [1995], miara zwana niejednoznacznością. W klasyfikacji natomiast częściej stosowana jest zero-jedynkowa funkcja straty. Miara niejednoznaczności przybiera wtedy następującą postać:

$$a_k(\mathbf{x}_n) = \begin{cases} 0, & \text{gdy } \hat{y}_k = \hat{y}_m, \\ 1 & \text{w przypadku przeciwnym,} \end{cases} \tag{7}$$

gdzie: $\hat{y}_k$ – wartość teoretyczna k-tego modelu.

Niejednoznaczność wszystkich pojedynczych modeli jest określana jako:

$$\bar{A} = \frac{1}{K}\sum_{k=1}^{K} A_k , \tag{8}$$

gdzie A_k to niejednoznaczność k-tego pojedynczego modelu dana wzorem:

$$A_k = \frac{1}{N}\sum_{n=1}^{N} a_k(\mathbf{x}_n) .$$

Miara zgodności modeli (*interrater agreement measure*)

Miara ta została zaproponowana przez J.L Fleissa [1981]:

$$IA = 1 - \frac{\sum_{n=1}^{N} m(\mathbf{x}_n)\left(K - m(\mathbf{x}_n)\right)}{NK(K-1)\bar{\varepsilon}(1-\bar{\varepsilon})}, \tag{9}$$

gdzie: $\bar{\varepsilon}$ to przeciętny błąd pojedynczych modeli, dany jako:

$$\bar{\varepsilon} = \frac{1}{K}\sum_{k=1}^{K} \varepsilon_k , \tag{10}$$

natomiast ε_k to błąd pojedynczego modelu, liczony jako:

$$\varepsilon_k = \frac{1}{N}\sum_{n=1}^{N} I\left\{y_n \neq \hat{y}_k\right\}. \tag{11}$$

Uogólniona miara zróżnicowania (*generalized diversity measure*)

Miara ta została zdefiniowana przez D. Partridge'a i W. Krzanowskiego [1997]. Niech Y oznacza zmienną losową przybierającą wartości równe odsetkowi pojedynczych modeli, które błędnie klasyfikują losowo wybraną obserwację $\mathbf{x}$. Przez p_k oznaczmy prawdopodobieństwo tego, że: $Y = \frac{k}{K}$, a przez $p(k)$ prawdopodobieństwo tego, że k losowo wybranych modeli błędnie zaklasyfikuje losowo wybraną obserwację $\mathbf{x}$. Wtedy:

$$p(1) = \sum_{k=1}^{K} \frac{k}{K} p_k , \tag{12}$$

$$p(2) = \sum_{k=1}^{K} \frac{k}{K} \frac{(k-1)}{(K-1)} p_k \,. \tag{13}$$

D. Partridge i W. Krzanowski stwierdzili, że maksymalne zróżnicowanie jest obserwowane wtedy, gdy dla losowej pary modeli błędowi jednego modelu towarzyszy poprawna klasyfikacja drugiego ($p(2) = 0$); natomiast minimalne zróżnicowanie odzwierciedla sytuację, gdy błędowi jednego modelu towarzyszy błąd drugiego modelu ($p(2) = p(1)$). Bazując na tych założeniach, zaproponowali prostą miarę

$$GD = 1 - \frac{p(2)}{p(1)} \,. \tag{14}$$

Miara ta przybiera wartości z przedziału [0, 1], gdzie 0 oznacza minimalne zróżnicowanie.

Miara współwystępowania błędów (*coincident failure diversity*)

Na potrzeby jednoczesnego współwystępowania wiekszej liczby błędów D. Partridge i W. Krzanowski (1997) przedstawili następującą miarę:

$$CFD = \begin{cases} 0, & \text{gdy } p_0 = 1 \\ \dfrac{1}{1-p_0} \displaystyle\sum_{k=1}^{K} \frac{K-k}{K-1} p_k, & \text{gdy } p_0 < 1. \end{cases} \tag{15}$$

Osiąga ona wartość maksymalną ($CFD = 1$), gdy wszystkie błędy zostaną popełnione przez jeden model, i wartość minimalną ($CFD = 0$), gdy wszystkie modele jednocześnie dokonają poprawnej albo niepoprawnej klasyfikacji.

Do miar dodatnio skorelowanych z błędem modelu zagregowanego należy miara podobieństwa modeli *IA* oraz miara trudności *DI*. Pozostałe natomiast należą do grupy miar ujemnie skorelowanych z błędem modelu zagregowanego.

3. Badanie relacji między miarami zróżnicowania a jakością modelu zagregowanego

W dotychczasowych badaniach porównywano różne miary zróżnicowania, analizując ich korelację z różnymi charakterystykami jakości modelu zagregowanego, takimi jak np. błąd modelu zagregowanego, różnica między błędem modelu zagregowanego a przeciętnym błędem pojedynczych modeli.

Ujawnienie się jasnej relacji między miarami zróżnicowania jest o tyle ważne, że coraz częściej pojawiają się algorytmy, które taki czynnik jak zróżnicowanie modeli w sposób jawny biorą pod uwagę w procesie konstrukcji modelu zagregowanego. Przykładami takich procedur mogą być algorytmy zaproponowane przez P. Cunninghama i J. Carneya [2000], G. Zenobi i P. Cunninghama [2001] oraz Opitza i Shavlika [1996].

Jedna z technik, która – jak to zostało dowiedzione – dobrze się nadaje do budowy modeli zagregowanych, polega na losowym doborze cech do budowy pojedynczych modeli.

Celem badania jest próba określenia, czy zaproponowane globalne miary zróżnicowania nadają się do określenia takich parametrów budowy pojedynczych modeli, które wchodzą w skład modelu zagregowanego, jak: liczba cech losowo wybieranych czy też liczba obserwacji w zbiorze uczącym, w taki sposób, by uzyskać jak najlepszą jakość modelu globalnego.

W tym celu zbadano stopień skorelowania miar zróżnicowania z:

- błędem modelu zagregowanego liczonym na osobnym zbiorze testowym (ε_{agr}),
- ilorazem błędu modelu globalnego i przeciętnego błędu pojedynczych modeli ($\varepsilon_{agr} / \overline{\varepsilon}$),
- różnicą między przeciętnym błędem modeli a błędem modelu zagregowanego ($\overline{\varepsilon} - \varepsilon_{agr}$).

Do eksperymentów wybrane zostały trzy ogólnodostępne zbiory danych, których charakterystyka znajduje się w tab. 1.

Tabela 1. Charakterystyka zbiorów danych

Nazwa	Liczba obserwacji	Liczba zmiennych	Liczba klas
Vehicle	846	18	4
Glass	214	10	6
Pima	768	8	2

Źródło: opracowanie własne.

Pierwszy eksperyment miał na celu zbadanie, czy miary zróżnicowania mogą być pomocne w określaniu liczby cech losowo wybieranych do budowy pojedynczych modeli. W tym celu dla zbioru Vehicle dobierano od 4 do 15 cech, a dla zbiorów Glass i Pima – od 3 do 7 cech, budując za każdym razem po 25 modeli składowych.

W przypadku tych trzech rozpatrywanych zbiorów danych okazuje się jednak, że korelacja między miarami zróżnicowania a błędem modelu zagregowanego jest bardzo niewielka. Znacznie wyższa korelacja zachodzi natomiast między miarami zróżnicowania a ilorazem błędu modelu zagregowanego i przeciętnego błędu pojedynczych modeli, a jeszcze wyższa między miarami zróżnicowania a różnicą między błędem przeciętnym a błędem modelu zagregowanego. W celu potwierdzenia wyników zbadano także korelacje między samymi tylko miarami zróżnicowania, uzyskując bardzo wysokie jej wartości, co świadczy o tym, że wszystkie te miary wykazują bardzo podobną wrażliwość na zmiany liczby cech.

Tabela 2. Miary zróżnicowania a liczba cech losowo dobieranych do budowy pojedynczych modeli

VEHICLE								
Korelacja między miarami zróżnicowania a różnymi charakterystykami jakości modelu zagregowanego								
Mierniki jakości modelu	Miara zróżnicowania							
	EN	LEN	DI	KW	AMB	IA	GD	CFD
ε_{agr}	-0,02	0,04	0,05	0,00	-0,03	0,02	-0,11	0,00
$\varepsilon_{agr} / \overline{\varepsilon}$	-0,83	-0,88	0,83	-0,88	-0,89	0,87	-0,85	-0,83
$\overline{\varepsilon} - \varepsilon_{agr}$	0,86	0,90	-0,83	0,89	0,90	-0,88	0,85	0,84

Korelacja między miarami zróżnicowania							
	lEN	DI	KW	AMB	IA	GD	CFD
EN	0,98	-0,94	1,00	0,99	-1,00	0,98	0,93
lEN		-0,93	0,99	0,98	-0,99	0,96	0,95
DI			-0,95	-0,93	0,96	-0,96	-0,98
KW				0,99	-1,00	0,98	0,95
AMB					-0,99	0,97	0,92
IA						-0,99	-0,95
GD							0,94

GLASS								
Korelacja między miarami zróżnicowania a różnymi charakterystykami jakości modelu zagregowanego								
Mierniki jakości modelu	Miara zróżnicowania							
	EN	LEN	DI	KW	AMB	IA	GD	CFD
ε_{agr}	-0,31	-0,18	0,22	-0,26	-0,16	0,37	-0,48	-0,25
$\varepsilon_{agr} / \overline{\varepsilon}$	-0,81	-0,74	0,73	-0,78	-0,70	0,81	-0,80	-0,73
$\overline{\varepsilon} - \varepsilon_{agr}$	0,61	0,66	-0,65	0,63	0,60	-0,58	0,53	0,60

Korelacja między miarami zróżnicowania							
	lEN	DI	KW	AMB	IA	GD	CFD
EN	0,97	-0,96	0,99	0,96	-0,98	0,93	0,93
lEN		-0,95	0,98	0,97	-0,94	0,85	0,93
DI			-0,96	-0,93	0,95	-0,90	-0,96
KW				0,96	-0,98	0,91	0,94
AMB					-0,93	0,85	0,89
IA						-0,98	-0,93
GD							0,87

PIMA								
Korelacja między miarami zróżnicowania a różnymi charakterystykami jakości modelu zagregowanego								
Mierniki jakości modelu	Miara zróżnicowania							
	EN	LEN	DI	KW	AMB	IA	GD	CFD
ε_{agr}	0,35	0,28	-0,11	0,31	0,35	-0,24	0,18	0,12
$\varepsilon_{agr} / \overline{\varepsilon}$	-0,82	-0,85	0,87	-0,84	-0,82	0,86	-0,88	-0,87
$\overline{\varepsilon} - \varepsilon_{agr}$	0,87	0,90	-0,88	0,89	0,87	-0,90	0,91	0,89

Korelacja między miarami zróżnicowania							
	lEN	DI	KW	AMB	IA	GD	CFD
EN	0,99	-0,92	0,99	0,99	-0,99	0,98	0,93
lEN		-0,95	0,99	0,99	-0,99	0,99	0,96
DI			-0,94	-0,92	0,95	-0,96	-0,99
KW				0,99	-1,00	0,99	0,94
AMB					-0,99	0,98	0,93
IA						-0,99	-0,96
GD							0,96

Źródło: obliczenia własne.

W drugim eksperymencie, mającym na celu zbadanie relacji między miarami zróżnicowania a błędem modelu zagregowanego w kontekście liczby obserwacji w zbiorze uczącym, służącym jako podstawa budowy pojedynczych modeli, pominięto zbiór Glass ze względu na małą liczebność. Dla pozostałych dwóch zaś – przy założeniu, że losowana będzie połowa cech – pojedyncze modele budowano, wykorzystując od 30 do 90% obserwacji z pierwotnego zbioru, dodając za każdym razem kolejne 10%.

Tabela 3. Miary zróżnicowania a liczba obserwacji w zbiorze uczącym

VEHICLE								
Korelacja między miarami zróżnicowania a różnymi charakterystykami jakości modelu zagregowanego								
Miemiki jakości modelu	Miara zróżnicowania							
	EN	lEN	DI	KW	AMB	IA	GD	CFD
ε_{agr}	-0,19	-0,05	-0,08	-0,13	-0,10	0,21	-0,41	-0,06
$\varepsilon_{agr} / \overline{\varepsilon}$	-0,63	-0,64	0,59	-0,61	-0,61	0,56	-0,40	-0,53
$\overline{\varepsilon} - \varepsilon_{agr}$	0,63	0,66	-0,62	0,62	0,62	-0,56	0,37	0,55

Korelacja między miarami zróżnicowania							
	lEN	DI	KW	AMB	IA	GD	CFD
EN	0,94	-0,88	0,99	0,95	-0,99	0,91	0,83
lEN		-0,89	0,95	0,98	-0,93	0,81	0,78
DI			-0,92	-0,84	0,89	-0,75	-0,92
KW				0,94	-0,99	0,91	0,87
AMB					-0,93	0,83	0,72
IA						-0,95	-0,86
GD							0,77

PIMA								
Korelacja między miarami zróżnicowania a różnymi charakterystykami jakości modelu zagregowanego								
Miemiki jakości modelu	Miara zróżnicowania							
	EN	lEN	DI	KW	AMB	IA	GD	CFD
ε_{agr}	0,05	0,10	-0,18	0,08	0,05	0,05	-0,19	0,08
$\varepsilon_{agr} / \overline{\varepsilon}$	-0,55	-0,53	0,35	-0,55	-0,55	0,57	-0,57	-0,29
$\overline{\varepsilon} - \varepsilon_{agr}$	0,63	0,61	-0,42	0,62	0,63	-0,63	0,59	0,35

Korelacja między miarami zróżnicowania							
	lEN	DI	KW	AMB	IA	GD	CFD
EN	0,95	-0,69	0,98	0,99	-0,96	0,88	0,52
lEN		-0,85	0,99	0,95	-0,97	0,88	0,72
DI			-0,79	-0,69	0,75	-0,67	-0,94
KW				0,98	-0,97	0,89	0,65
AMB					-0,96	0,88	0,52
IA						-0,97	-0,64
GD							0,60

Źródło: obliczenia własne.

Na podstawie uzyskanych wyników widać wyraźnie, że w porównaniu z błędem modelu zagregowanego znacznie wyższą korelację odnotowujemy między miarami zróżnicowania a pozostałymi dwiema charakterystykami jakości modelu zagregowanego, tj. $\varepsilon_{agr} / \overline{\varepsilon}$ i $\overline{\varepsilon} - \varepsilon_{agr}$.

Również korelacje między samymi tylko miarami zróżnicowania kształtują się na wysokim poziomie, więc trudno wskazać, która z nich jest najlepsza.

4. Wnioski

Można stwierdzić, że:

- zachodzi słaba korelacja między wszystkimi miarami zróżnicowania a błędem modelu zagregowanego. Wynika stąd zatem, że na podstawie wartości tych miar, obliczanych przy różnych wartościach parametrów w algorytmach budowy pojedynczych modeli, nie można wnioskować o błędzie modelu zagregowanego;

- silniejszą korelację zauważa się w przypadku badania siły związku między miarami zróżnicowania a innymi charakterystykami jakości modelu zagregowanego, takimi jak np. iloraz błędu modelu zagregowanego i przeciętnego błędu pojedynczych modeli;
- wysokie wartości współczynników korelacji między samymi tylko miarami zróżnicowania świadczą o tym, że wszystkie te miary wykazują podobne tendencje i w związku z tym trudno jest wskazać, która z nich jest najlepsza i powinna być rekomendowana jako ta, która pomoże zbudować najlepszy model zagregowany.

Literatura

Blake C., Keogh E., Merz C.J. (1998), *UCI Repository of Machine Learning Databases*, Department of Information and Computer Science, University of California, Irvine.

Cunningham P., Carney J. (2000), *Diversity Versus Quality in Classification Ensembles Based on Feature Selection*, Technical Report TCD-CS-2000-02, Department of Computer Science, Trinity College, Dublin.

Fleiss J.L. (1981), *Statistical Methods for Rates and Proportions*, John Willey and Sons.

Hansen L.K., Salamon P. (1990), *Neural Networks Ensembles*, „IEEE Transactions on Pattern Analysis and Machine Intelligence" 12 (10).

Kohavi R., Wolpert D.H. (1996), *Bias Plus Variance Decomposition for Zero-one Loss Function*, „Proceedings of the 12th International Conference on Machine Learning", Morgan Kaufmann, Tahoe City.

Kuncheva L.I., Whitaker C.J. (2001), *Feature Subsets for Classifier Combination: an Enumerative Experiment*, „Proceedings of the 2nd International Workshop on Multiple Classifier Systems", Springer-Verlag, Cambridge, UK.

Kuncheva L.I., Whitaker C.J. (2003), *Measures of Diversity in Classifier Ensembles*, „Machine Learning" 51.

Krogh A., Vedelsby J. (1995), *Neural Network Ensembles, Cross Validation, and Active Learning*, NIPS 7.

Opitz D., Shavlik J. (1996), *Generating Accurate and Diverse Members of a Neural Network Ensemble*, „Advances in Neural Information Processing Systems", MIT Press, Denver.

Partridge D., Krzanowski W. (1997), *Software Diversity: Practical Statistics for Its Measurement and Exploitation*, „Information and Software Technology" 39.

Ruta D., Gabrys B. (2002a), *A Theoretical Analysis of the Limits of Majority Voting Errors for Multiple Classifier Systems*, „Pattern Analysis and Applications" 5 (4).

Ruta D., Gabrys B. (2002b), *Set Analysis of Coincident Errors and Its Applications for Combining Classifiers*, „Pattern Recognition and String Maching, Combinatorial Optimisation", 13, Kluwer Academic Publishers.
Zenobi G., Cunningham P. (2001), *Using Diversity in Preparing Ensembles of Classifiers Based on Different Feature Subsets to Minimise Generalisation Error*, „Proceedings of the 12th European Conference on Machine Learning".

ANALYSIS OF RELATIONSHIP BETWEEN DIVERSITY MEASURES AND ERROR OF AGGREGATED CLASSIFIER

Summary

The main aim of ensembles is improvement of their classification and prediction accuracy. One of the elements required for accurate prediction when using an ensemble is recognised to be error „diversity".

The paper presents different diversity measures and disscuss their main properties. It shows also the influence of different factors (eg. number of single models, the training set size) on the diversity of base classifiers of the ensemble.