Zastosowanie sieci neuronowych w problemie klasyfikacji wielokategorialnej

Adam Żychowski

Definicja problemu

$X \subseteq \mathbb{R}^d$ - zbiór obiektów

$Y = \{y_1, y_2, \ldots, y_Q\}$ - zbiór kategorii

Cel: funkcja $h : X \to 2^Y$

Każdy z obiektów może należeć do więcej niż jednej kategorii.

Alternatywna definicja

$X \subseteq \mathbb{R}^d$ - zbiór obiektów

$Y = \{y_1, y_2, \ldots, y_Q\}$ - zbiór kategorii

Cel: funkcja $f : X \times Y \to \mathbb{R}$ taka, że

$$\forall x_p \in X, \forall y_1 \in Y_p, y_2 \notin Y_p \qquad f(x_p, y_1) > f(x_p, y_2)$$

Zastosowania

- kategoryzacja tekstów (tagowanie artykułów, e-maili)
- multimedia (obrazy, filmy, muzyka)
- biologia (odkrywanie funkcji genomów, rozpoznawanie chorób na podstawie symptomów)

Dotychczasowe rozwiązania

Podział ze względu na badany poziom korelacji:

- pierwszego rzędu (traktowanie każdej kategorii oddzielnie)
- drugiego rzędu (badanie korelacji między parami kategorii)
- wyższych rzędów (uwzględnianie korelacji między zbiorami kategorii)

Dotychczasowe rozwiązania

- sprowadzenie problemu do klasycznego problemu klasyfikacji (Binary Relevance, Classifier Chaining)
- modyfikacja istniejącego algorytmu i dostosowanie go do problemu klasyfikacji wielokryterialnej (ML-kNN, Predictive clustering trees)
- metody hybrydowe

Sieci neuronowe

Pierwsza propozycja zastosowania sieci neuronowych w problemie klasyfikacji wielokryterialnej:

Zhang, M. L., and Zhou, Z. H. Multilabel neural networks with applications to functional genomics and text categorization. IEEE transactions on Knowledge and Data Engineering, 18(10), 1338-1351. 2006.

- perceptron wielowarstwowy
- uczenie metodą propagacji wstecznej błędu
- jedna warstwa ukryta (40 neuronów)
- wejście: cechy obiektu
- *Q* wyjść: każde odpowiadające jednej kategorii

Błąd uczenia wersja 1 (BP-MLL)

$$E_1 = \sum_{p=1}^{m} \frac{\sum_{(r,s)\in Y_p \times \overline{Y}_p} e^{-\left(c_r^p - c_s^p\right)}}{\left|Y_p\right|\left|\overline{Y}_p\right|}$$

$$h(x_p) = \left\{q \in Y : c_q(x_p) > t(x_p)\right\}, \quad c_q(x_p) = c_q^p$$

m - liczba obiektów w zbiorze uczacym

$Y_p \subseteq Y$ - zbiór kategorii, do których należy p-ty obiekt

$\overline{Y_p}$ - zbiór kategorii, do których nie należy p-ty obiekt ($\overline{Y_p} = Y \setminus Y_p$)

c_q^p - aktualna wartość wyjścia neuronu odpowiadajacego q-tej kategorii dla p-tego obiektu

$t(x_p)$ - próg

Modyfikacja błędu uczenia:

Grodzicki Rafał, Mańdziuk Jacek, and Wang Lipo. „Improved multilabel classification with neural networks.” *International Conference on Parallel Problem Solving from Nature*. Springer Berlin Heidelberg, 2008.

Błąd uczenia wersja 2

$$E_2 = \sum_{p=1}^{m} \frac{\sum_{(r,s)\in Y_p\times\bar{Y}_p} e^{-\left(c_r^p - c_s^p\right)} + \sum_{r\in Y_p} e^{-\left(c_r^p - c_Q^p\right)} + \sum_{s\in\bar{Y}_p} e^{-\left(c_Q^p - c_s^p\right)}}{\left|Y_p\right|\left|\bar{Y}_p\right| + \left|Y_p\right| + \left|\bar{Y}_p\right|}$$

$$h\left(x_p\right) = \left\{q \in Y : c_q\left(x_p\right) > c_Q\left(x_p\right)\right\}, \quad c_q\left(x_p\right) = c_q^p$$

$c_Q^{\bar{p}}$ - aktualna wartość wyjścia dodatkowego neuronu
- wartość progu

Błąd uczenia wersja 3

$$E_3 = \sum_{p=1}^{m} \frac{\sum_{(r,s)\in Y_p\times\bar{Y}_p} e^{-\left(c_{2r}^p - c_{2s}^p\right)} + \sum_{r\in Y_p} e^{-\left(c_{2r}^p - c_{2r+1}^p\right)} + \sum_{s\in\bar{Y}_p} e^{-\left(c_{2s+1}^p - c_{2s}^p\right)}}{\left|Y_p\right|\left|\bar{Y}_p\right| + \left|Y_p\right| + \left|\bar{Y}_p\right|}$$

$$h\left(x_p\right) = \left\{q \in Y : c_{2q}\left(x_p\right) > c_{2q+1}\left(x_p\right)\right\}, \quad c_q\left(x_p\right) = c_q^p$$

Błąd uczenia wersja 4 oraz 5 (CART-M)

$$E_4 = \sum_{p=1}^{m} \frac{\sum_{(r,s)\in Y_p\times\bar{Y}_p} e^{-\left(c_{2r}^p - c_{2s}^p\right)} + \sum_{r\in Y_p}\sum_{t\in Y_p} e^{-\left(c_{2r}^p - c_{2t+1}^p\right)} + \sum_{s\in \bar{Y}_p}\sum_{t\in \bar{Y}_p} e^{-\left(c_{2t+1}^p - c_{2s}^p\right)}}{\left|Y_p\right|\left|\bar{Y}_p\right| + \left|Y_p\right|^2 + \left|\bar{Y}_p\right|^2}$$

$$h\left(x_p\right) = \left\{q \in Y : c_{2q}\left(x_p\right) > c_{2q+1}\left(x_p\right)\right\}, \quad c_q\left(x_p\right) = c_q^p$$

$$E_5 = \sum_{p=1}^{m} \frac{\sum_{(r,s)\in Y_p\times\bar{Y}_p} \left(e^{-\left(c_{2r}^p - c_{2s}^p\right)} + e^{-\left(c_{2s+1}^p - c_{2r+1}^p\right)}\right) + \sum_{r\in Y_p}\sum_{t\in Y_p} e^{-\left(c_{2r}^p - c_{2t+1}^p\right)} + \sum_{s\in \bar{Y}_p}\sum_{t\in \bar{Y}_p} e^{-\left(c_{2t+1}^p - c_{2s}^p\right)}}{2\left|Y_p\right|\left|\bar{Y}_p\right| + \left|Y_p\right|^2 + \left|\bar{Y}_p\right|^2}$$

$$h\left(x_p\right) = \left\{q \in Y : c_{2q}\left(x_p\right) > c_{2q+1}\left(x_p\right)\right\}, \quad c_q\left(x_p\right) = c_q^p$$

Błąd uczenia wersja 6 (TCART-M)

$$E_6 = \sum_{p=1}^{m} \frac{\sum_{(r,s)\in Y_p\times\bar{Y}_p}\left(e^{-\left(c_{2r}^p - c_{2s}^p\right)} + \frac{e^{-\left(c_{2s+1}^p - c_{2r+1}^p\right)}}{D}\right) + \sum_{r\in Y_p}\sum_{t\in Y_p}\frac{e^{-\left(c_{2r}^p - c_{2t+1}^p\right)}}{D} + \sum_{s\in \bar{Y}_p}\sum_{t\in \bar{Y}_p}\frac{e^{-\left(c_{2t+1}^p - c_{2s}^p\right)}}{D}}{2\left|Y_p\right|\left|\bar{Y}_p\right| + \left|Y_p\right|^2 + \left|\bar{Y}_p\right|^2}$$

Mańdziuk Jacek, Żychowski Adam, and Wang Lipo. „A TCART-M - Tuned CARTesian-based Error Function for Multilabel Classification with the MLP.” IJCNN 2017.

Benchmarki

yeast – zbiór genomów i ich funkcji

scene – obrazy krajobrazów

emotions – próbki dźwięków (piosenek), którym przypisywane są emocje

enron – e-maile 150 pracowników firmy Enron Corporation

medical – opisy symptomów choroby pacjentów

Name	Domain	Instances	Attributes	Labels	Card.
yeast	biology	2417	103	14	4.24
scene	multimedia	2407	294	6	1.07
emotions	multimedia	593	72	6	1.87
enron	text	1702	1001	53	3.38
medical	text	978	1449	45	1.25

Miary błędu rozwiązania

- *example-based measures* – różnica w kategoriach przypisanych przez metodę a prawidłowych obliczana jest dla każdego obiektu oddzielnie, a następnie uśredniana (*Hamming loss*, *accuracy*, *precision*, *recall*, *F1 score*, *subset accuracy*)

Przykład: $$precision(h) = \frac{1}{N}\sum_{i=1}^{n}\frac{|h(x_i) \cap Y_i|}{|Y_i|}$$

- *label-based measures* – obliczana jest różnica dla wszystkich obiektów dla danej kategorii, a potem wartość jest uśredniana względem wszystkich kategorii (*micro-precision*, *micro-recall*, *micro-F_1*, *macro-precision*, *macro-recall*, *macro-F_1*)

Przykład: $$macro\_precision = \frac{1}{Q}\sum_{j=1}^{Q}\frac{tp_j}{tp_j + fp_j}$$

tp_j - liczba prawidłowo przypisanych obiektów do kategorii (true positives)
fp_j - liczba nieprawidłowo przypisanych obiektów do kategorii (false positives)

Miary błędu rozwiązania

- *ranking-based measures* – miary wyliczane na podstawie kolejności (od najbardziej prawdopodobnej do najmniej), w jakiej metoda przypisała kategorie (*average precision*, *one-error*, *coverage*, *ranking loss*)

Przykład: $$rloss(f) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{|Y_i||\bar{Y}_i|}|\{(y', y'') \mid f(\boldsymbol{x}_i, y') \leq f(\boldsymbol{x}_i, y''), (y', y'') \in Y_i \times \bar{Y}_i)\}|$$

Testy metody TCART-M

$$D_{cand} = \{0.25, 0.5, 0.75, 1.0, 1.25, 1.5, 1.75, 2, 2.5, 3, 3.5, 4, 5, 6, 8, 10\}$$

Dobór *D* metodą Nested Cross Validation.

Dwa podejścia:

- TCART-M_i – dla każdej z miar błędu oddzielnie dobierany jest parametr *D*
- TCART-M_g – próba znalezienia uniwersalnej wartości *D*, tworzenie rankingu po wszystkich miarach błędu

Wyniki dla benchmarku *emotions*

	BP-MLL	CART-M	TCART-M*g*	TCART-M*i*
Hamming Loss	0.203	0.201	0.187	**0.186**
Accuracy	0.570	0.547	0.579	**0.580**
Precision	0.652	0.666	0.682	**0.683**
Recall	**0.730**	0.663	0.700	0.703
Subset Accuracy	0.299	0.291	0.331	**0.332**
F1 score	0.660	0.632	0.660	**0.663**
Micro-precision	0.657	0.685	**0.701**	0.695
Macro-precision	0.657	0.685	**0.700**	0.696
Micro-recall	**0.728**	0.658	0.698	0.706
Macro-recall	**0.714**	0.650	0.686	0.703
Micro-F1	0.690	0.671	**0.699**	0.693
Macro-F1	0.676	0.658	0.685	**0.690**
Ranking Loss	0.160	0.160	0.145	**0.142**
OneError	0.290	0.275	**0.251**	0.253
Coverage	1.748	1.766	1.694	**1.658**
Average Precision	0.799	0.803	**0.818**	0.812

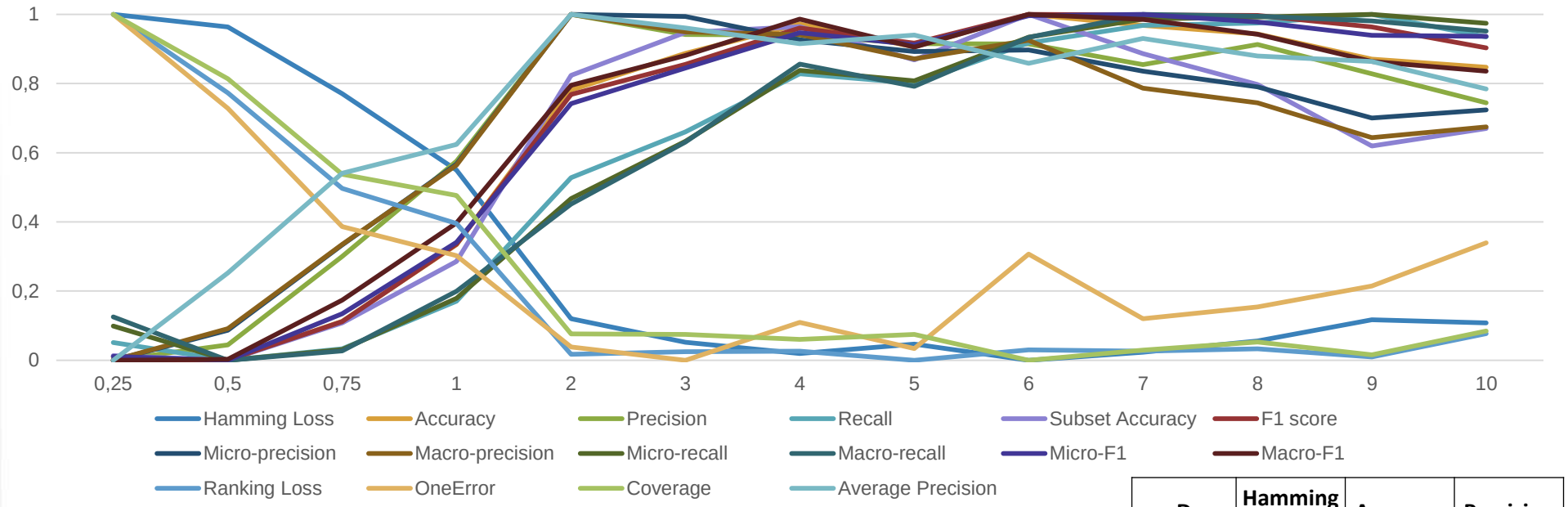
Wyniki dla benchmarku *yeast*

	BP-MLL	CART-M	TCART-M*g*	TCART-M*i*
Hamming Loss	0.217	**0.197**	0.198	0.198
Accuracy	**0.535**	0.510	0.510	0.509
Precision	0.637	0.700	**0.706**	0.704
Recall	**0.711**	0.595	0.596	0.601
Subset Accuracy	0.143	**0.168**	0.160	0.161
F1 score	0.017	0.600	0.618	**0.621**
Micro-precision	0.626	**0.709**	**0.709**	0.706
Macro-precision	0.464	**0.544**	0.533	0.541
Micro-recall	**0.702**	0.592	0.590	0.590
Macro-recall	**0.468**	0.374	0.359	0.369
Micro-F1	**0.661**	0.645	0.644	0.643
Macro-F1	**0.441**	0.405	0.383	0.406
Ranking Loss	0.174	**0.164**	0.165	**0.164**
OneError	0.237	0.227	0.225	**0.224**
Coverage	6.441	6.285	6.305	**6.265**
Average Precision	0.754	**0.767**	0.766	**0.767**

CART-M vs TCART-M

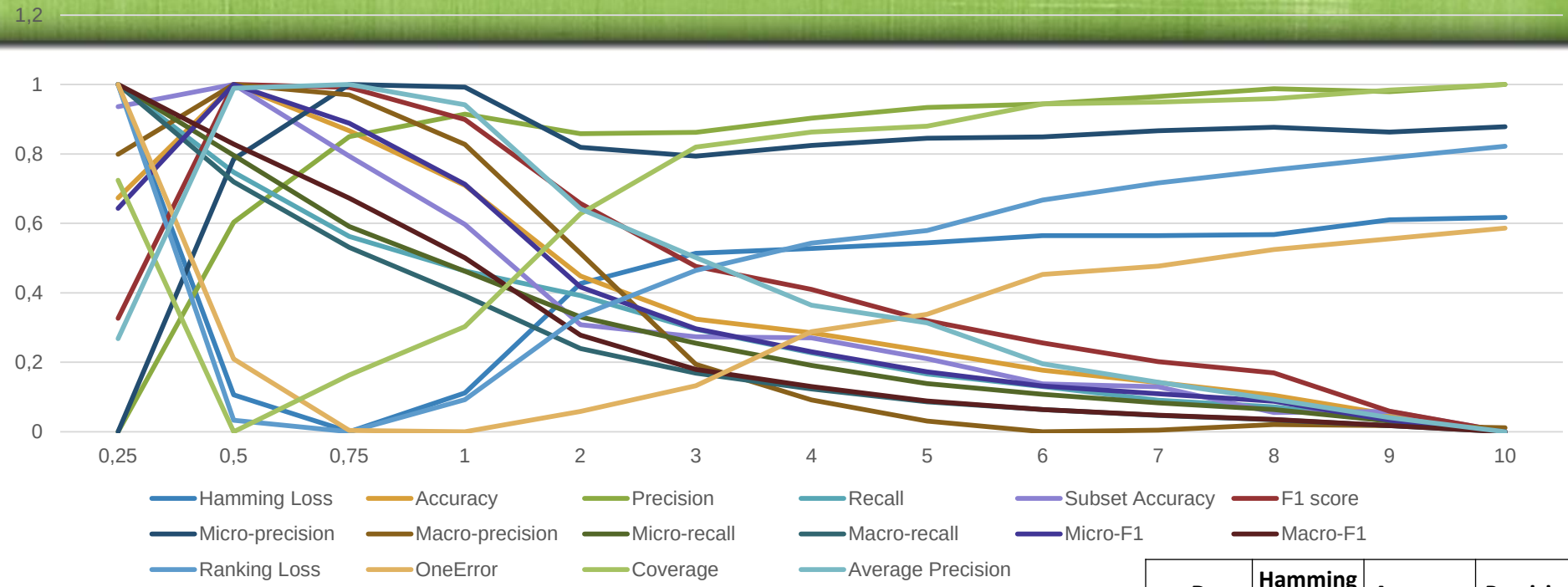
	TCART-M*g*		TCART-M*i*	
Benchmark	Mean	D	Mean	D
emotions	16 (16)	6.7	16 (16)	6.1
yeast	4 (2)	2.0	7 (4)	2.5
scene	12 (9)	5.1	12 (11)	6.3
enron	11 (8)	4.7	13 (9)	4.5
medical	12 (7)	2.0	12 (7)	2.1

Porównanie wartości *D* dla benchmarku *emotions*



D	Hamming Loss	Accuracy	Precision
0,25	0,204	0,546	0,656
0,5	0,203	0,545	0,658
0,75	0,199	0,549	0,666
1	0,194	0,558	0,674
2	0,186	0,574	**0,687**
3	0,184	0,578	0,686
4	**0,183**	0,582	0,686
5	0,184	0,579	0,685
6	**0,183**	**0,583**	0,685
7	0,184	0,581	0,683
8	0,184	0,581	0,685
9	0,185	0,578	0,682
10	0,185	0,577	0,679

Porównanie wartości *D* dla benchmarku *yeast*



D	Hamming Loss	Accuracy	Precision
0,25	0,205	0,504	0,674
0,5	0,197	0,510	0,699
0,75	0,196	0,508	0,709
1	0,197	0,505	0,712
2	0,200	0,501	0,710
3	0,201	0,498	0,710
4	0,201	0,498	0,711
5	0,201	0,497	0,713
6	0,201	0,496	0,713
7	0,201	0,495	0,714
8	0,201	0,495	0,715
9	0,201	0,494	0,714
10	0,202	0,493	0,715

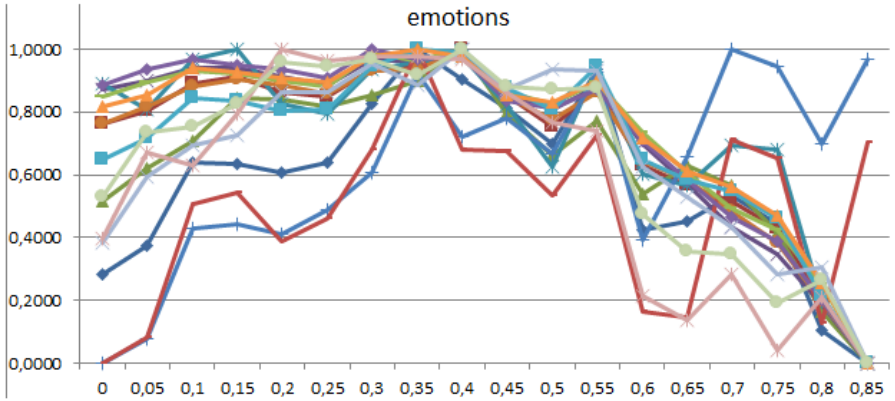
Porównanie z innymi metodami benchmarku *emotions*

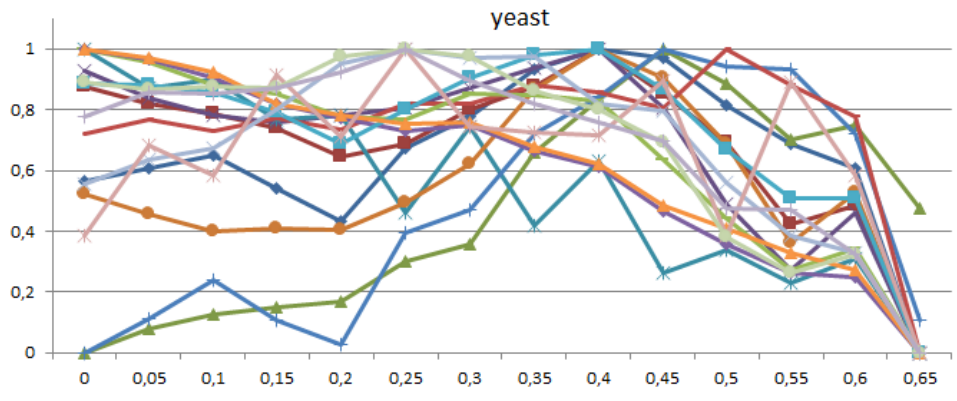
	BP-MLL	CART-M	TCART-M*g*	TCART-M*i*	BR	CC	CLR	QWML	HOMER	ML-C4.5	PCT	ML-kNN	RAkEL	ECC	RFML-C4.5	RF-PCT
Hamming Loss	0.203	0.201	0.187	**0.186**	0.257	0.256	0.257	0.254	0.361	0.247	0.267	0.294	0.282	0.281	0.198	0.189
Accuracy	0.570	0.547	0.579	**0.580**	0.361	0.356	0.361	0.373	0.471	0.536	0.448	0.319	0.419	0.432	0.488	0.519
Precision	0.652	0.666	0.682	**0.683**	0.550	0.551	0.538	0.548	0.509	0.606	0.577	0.502	0.564	0.580	0.625	0.644
Recall	0.730	0.663	0.700	0.704	0.409	0.397	0.410	0.429	**0.775**	0.703	0.534	0.377	0.491	0.533	0.545	0.582
Subset Accuracy	0.299	0.291	0.331	**0.332**	0.129	0.124	0.144	0.149	0.163	0.277	0.223	0.084	0.208	0.168	0.272	0.307
F1 score	0.660	0.632	0.660	**0.663**	0.469	0.461	0.465	0.481	0.614	0.651	0.554	0.431	0.525	0.556	0.583	0.611
Micro-precision	0.657	0.685	0.701	0.695	0.684	0.698	0.685	0.680	0.471	0.607	0.607	0.584	0.586	0.579	**0.783**	**0.783**
Macro-precision	0.657	0.685	0.700	0.697	0.721	0.581	0.677	0.660	0.464	0.602	0.628	0.518	0.547	0.531	**0.828**	0.802
Micro-recall	0.728	0.658	0.698	0.706	0.406	0.393	0.409	0.431	**0.782**	0.712	0.539	0.376	0.489	0.531	0.551	0.589
Macro-recall	0.714	0.650	0.686	0.703	0.378	0.364	0.381	0.398	**0.775**	0.702	0.533	0.334	0.462	0.508	0.532	0.569
Micro-F1	0.690	0.671	**0.699**	0.693	0.509	0.503	0.512	0.528	0.588	0.655	0.571	0.457	0.533	0.554	0.647	0.672
Macro-F1	0.676	0.658	0.684	**0.690**	0.440	0.420	0.443	0.458	0.570	0.630	0.568	0.385	0.488	0.500	0.620	0.650
Ranking Loss	0.160	0.160	0.145	**0.142**	0.246	0.245	0.264	0.331	0.297	0.210	0.270	0.283	0.281	0.310	0.153	0.151
OneError	0.290	0.275	**0.251**	0.253	0.386	0.376	0.391	0.391	0.411	0.347	0.386	0.406	0.396	0.426	0.277	0.262
Coverage	1.748	1.766	1.694	**1.658**	2.307	2.317	2.376	2.807	2.634	2.069	2.356	2.490	2.465	2.619	1.801	1.827
Average Precision	0.799	0.803	**0.818**	0.812	0.721	0.724	0.718	0.679	0.698	0.759	0.713	0.694	0.713	0.687	0.812	0.812

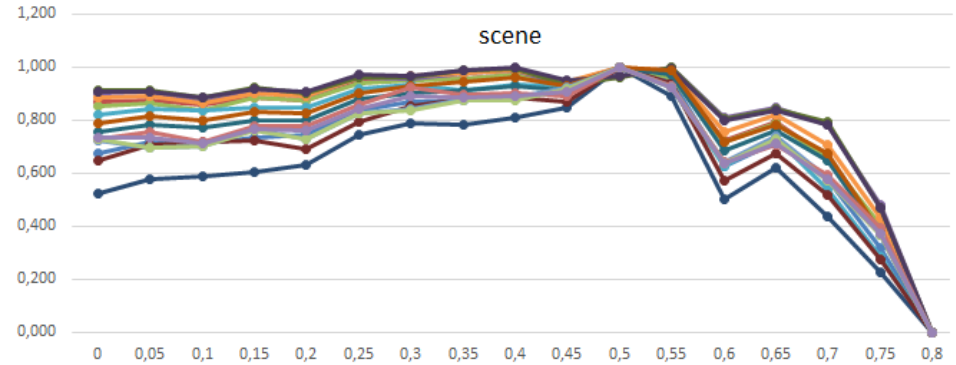
Porównanie z innymi metodami

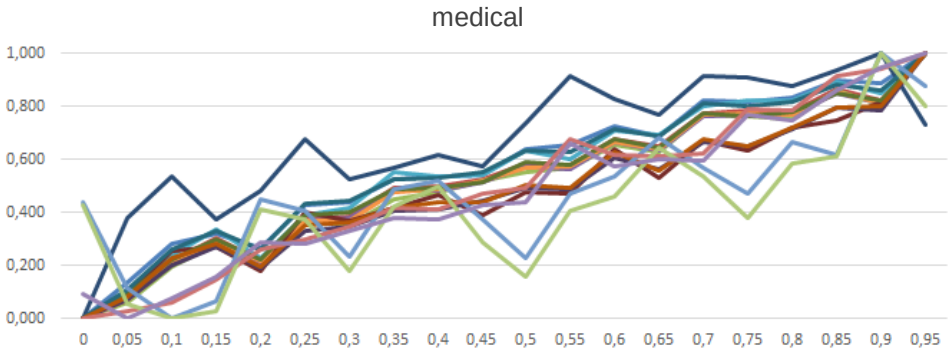
emotions	yeast	scene	enron	medical	sum
TCART-M*i* (34)	CLR (72)	BR (78)	**TCART-M*i*** (74)	QWML (76)	**TCART-M*i*** (942)
TCART-M*g* (41)	BR (99)	CC (79)	BR (84)	HOMER (83)	**TCART-M*g*** (1026)
BP-MLL (72)	HOMER (110)	RAkEL (79)	**TCART-M*g*** (86)	CLR (90)	CLR (1056)
RF-PCT (72)	**TCART-M*i*** (115)	ECC (90)	CLR (89)	ML-C4,5 (94)	BR (1152)
CART-M (79)	CC (115)	CLR (92)	CART-M (106)	RF-PCT (106)	HOMER (1166)
RFML-C4,5 (90)	CART-M (117)	**TCART-M*i*** (103)	RF-PCT (110)	RAkEL (120)	RF-PCT (1166)
ML-C4,5 (102)	QWML (123)	**TCART-M*g*** (106)	HOMER (119)	CC (124)	CART-M (1176)
PCT (154)	**TCART-M*g*** (132)	HOMER (107)	ECC (127)	BR (137)	CC (1306)
HOMER (164)	ECC (132)	CART-M (124)	RFML-C4,5 (137)	ECC (137)	RAkEL (1338)
BR (178)	RAkEL (135)	QWML (133)	CC (146)	**TCART-M*i*** (145)	ECC (1350)
RAkEL (184)	BP-MLL (138)	RF-PCT (155)	RAkEL (151)	**TCART-M*g*** (148)	QWML (1380)
CLR (185)	RF-PCT (140)	ML-k NN (182)	QWML (165)	ML-k NN (149)	RFML-C4,5 (1536)
CC (189)	ML-k NN (145)	RFML-C4,5 (185)	ML-C4,5 (171)	CART-M (162)	ML-C4,5 (1554)
ECC (189)	RFML-C4,5 (180)	BP-MLL (208)	BP-MLL (174)	RFML-C4,5 (176)	BP-MLL (1586)
QWML (193)	ML-C4,5 (194)	ML-C4,5 (216)	ML-k NN (195)	BP-MLL (201)	ML-k NN (1826)
ML-k NN (242)	PCT (218)	PCT (233)	PCT (232)	PCT (215)	PCT (2104)

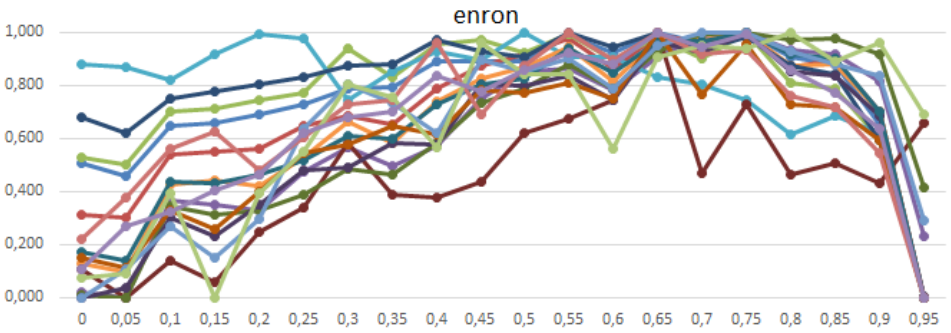
Redukcja danych wejściowych











Dalsze eksperymenty

- sprawdzenie innych benchmarków
- „sprytniejszy” dobór wartości parametru *D (np. algorytm genetyczny)*
- *rozważenie różnych dzielników dla różnych składników sumy*
- *wykorzystanie informacji z nauczonej sieci*

Literatura

1. Zhang, M. L., and Zhou, Z. H. A review on multi-label learning algorithms. *IEEE transactions on knowledge and data engineering*, *26*(8), 1819-1837. 2014.

2. Madjarov, G., Kocev, D., Gjorgjevikj, D., and Džeroski, S. *An extensive experimental comparison of methods for multi-label learning*. Pattern Recognition, 45(9), 3084-3104. 2012.

3. Zhang, M. L., and Zhou, Z. H. *Multilabel neural networks with applications to functional genomics and text categorization*. IEEE transactions on Knowledge and Data Engineering, *18*(10), 1338-1351. 2006.

4. Grodzicki, R., Mańdziuk, J., and Wang, L. *Improved multilabel classification with neural networks*. International Conference on Parallel Problem Solving from Nature (pp. 409-416). Springer Berlin Heidelberg. 2008.

5. Tsoumakas G., Spyromitros-Xioufis E., Vilcek J., and Vlahavas I. *Mulan: A Java Library for Multi-Label Learning*, Journal of Machine Learning Research, 12, pp. 2411-2414. 2011. http://mulan.sourceforge.net