

Testowanie

Michał Mazek

Andrzej Ratajczak

Jan Grzegorowski

26 stycznia 2014

1 Raport z testów

W raporcie znajduje się zestawienie statystyk ze wszystkich przeprowadzonych testów. W każdej tabeli znajdują się szczegółowe dane dotyczące wybranego klasyfikatora, natomiast na dole tabeli widoczne są sumaryczne statystyki. Niestety wyniki przedstawione w tabelach różnią się od tych opisywanych w raporcie ze względu na inny zbiór testowy pytań i odpowiedzi zawierający 25, a nie 30 przykładów.

Kolumna "dokładność" przedstawia ogólną dokładność klasyfikacji (dla pojedynczego tematu oraz dla całego zbioru), tj. odsetek odpowiedzi zaklasyfikowanych zgodnie z klasyfikacją zadaną w zbiorze testowym. W kolumnie "poprawna" znajduje się odsetek prawidłowo zaklasyfikowanych odpowiedzi *poprawnych*. Kolumna "najlepsza" przedstawia natomiast zbieżność wyboru najlepszej odpowiedzi z odpowiedzią oznaczoną jako poprawną. Średnia wartość tej ostatniej oznacza rzeczywistą skuteczność wyboru właściwej odpowiedzi przez aplikację.

Można zauważyć, że odpowiedzi poprawne bardzo często nie są jako takie klasyfikowane (niskie wartości w kolumnie "poprawna"). Wynika to z tego, że klasyfikatory probabilistyczne przyjmują wstępne prawdopodobieństwo dla każdej klasy proporcjonalne do częstości jej występowania w zbiorze trenującym; jako, że na każdy temat przypada tylko jedna odpowiedź oznaczona jako poprawna (tj. wybrana przez zakładającego temat), a odpowiedzi niepoprawnych może być dowolnie dużo, wstępne prawdopodobieństwo dla klasy "niepoprawne" jest dużo wyższe, niż dla klasy "poprawne". Potocznie można powiedzieć, że klasyfikatory są "wybredne". Mimo to wybór najlepszej odpowiedzi jako posiadającej najwyższe prawdopodobieństwo klasyfikacji jako poprawnej, niezależnie od tego, czy

została tak zaklasyfikowana, pozwala na wykorzystanie takiego "wybrednego" klasyfikatora, co potwierdza stosunkowo wysoka (w stosunku do dokładności samej klasyfikacji) skuteczność tej metody widoczna w niniejszych testach.

Przeprowadzenie dokładniejszych testów wykazało też, że wbrew początkowym spostrzeżeniom dokonanych na podstawie samej dokładności klasyfikacji, wykorzystanie słownika w zestawie cech wpływa korzystnie na wybór poprawnej odpowiedzi, zmniejszając tym samym "wybredność" klasyfikatora. Jego wykorzystanie jest jednak uzasadnione jedynie w połączeniu z klasyfikatorem bayesowskim.

2 Wyniki testów

Klasyfikator: NaiveBayes kontekst			
Nr. testu	Dokładność	Poprawna	Najlepsza
1	0.80	0	1.00
2	0.80	0	1.00
3	0.80	0	1.00
4	0.50	0	1.00
5	0.80	0	1.00
6	0.67	0	1.00
7	0.80	0	1.00
8	0.67	0	1.00
9	0.67	0	1.00
10	0.83	0	1.00
11	0.80	0	1.00
12	0.67	0	0.00
13	0.67	0	1.00
14	0.50	0	1.00
15	0.88	0	0.00
16	0.86	0	1.00
17	0.75	0	1.00
18	0.80	0	1.00
19	0.50	0	1.00
20	0.75	0	0.00
21	0.80	0	1.00
22	0.75	0	0.00
23	0.67	0	1.00
24	0.50	0	1.00
25	0.75	0	0.00
Razem	0.61	0.00	0.80

Klasyfikator: NaiveBayes słownik			
Nr. testu	Dokładność	Poprawna	Najlepsza
1	0.20	1	1.00
2	0.20	1	1.00
3	0.20	1	1.00
4	0.50	1	1.00
5	0.20	1	1.00
6	0.33	1	1.00
7	0.20	1	1.00
8	0.33	1	1.00
9	0.33	1	1.00
10	0.17	1	1.00
11	0.20	1	1.00
12	0.33	1	1.00
13	0.33	1	1.00
14	0.50	1	1.00
15	0.12	1	0.00
16	0.14	1	1.00
17	0.25	1	1.00
18	0.20	1	1.00
19	0.50	1	1.00
20	0.25	1	0.00
21	0.20	1	1.00
22	0.25	1	1.00
23	0.33	1	1.00
24	0.50	1	1.00
25	0.25	1	1.00
Razem	0.20	1.00	0.92

Klasyfikator: NaiveBayes razem			
Nr. testu	Dokładność	Poprawna	Najlepsza
1	0.20	1	1.00
2	0.20	1	1.00
3	0.20	1	1.00
4	0.50	1	1.00
5	0.20	1	1.00
6	0.33	1	1.00
7	0.20	1	1.00
8	0.33	1	1.00
9	0.33	1	1.00
10	0.17	1	1.00
11	0.20	1	1.00
12	0.33	1	1.00
13	0.33	1	1.00
14	0.50	1	1.00
15	0.12	1	0.00
16	0.29	1	1.00
17	0.25	1	1.00
18	0.20	1	1.00
19	0.50	1	1.00
20	0.25	1	0.00
21	0.20	1	1.00
22	0.25	1	1.00
23	0.33	1	1.00
24	0.50	1	1.00
25	0.25	1	1.00
Razem	0.20	1.00	0.92

Klasyfikator: NaiveBayes słowa kluczowe			
Nr. testu	Dokładność	Poprawna	Najlepsza
1	0.80	0	0.00
2	0.80	0	1.00
3	0.80	0	1.00
4	0.50	0	0.00
5	0.80	0	0.00
6	0.67	0	1.00
7	0.80	0	0.00
8	0.67	0	0.00
9	0.67	0	0.00
10	0.83	0	0.00
11	0.80	0	0.00
12	0.67	0	1.00
13	0.67	0	1.00
14	0.50	0	0.00
15	0.88	0	0.00
16	0.86	0	1.00
17	0.75	0	1.00
18	0.80	0	0.00
19	0.50	0	1.00
20	0.75	0	1.00
21	0.80	0	0.00
22	0.75	0	1.00
23	0.67	0	0.00
24	0.50	0	0.00
25	0.75	0	0.00
Razem	0.61	0.00	0.40

Klasyfikator: NaiveBayes bez słownika			
Nr. testu	Dokładność	Poprawna	Najlepsza
1	0.80	0	0.00
2	0.80	0	1.00
3	0.80	0	1.00
4	0.50	0	1.00
5	0.80	0	0.00
6	0.67	0	1.00
7	0.80	0	0.00
8	0.67	0	1.00
9	0.67	0	0.00
10	0.83	0	0.00
11	0.80	0	1.00
12	0.67	0	0.00
13	0.67	0	1.00
14	0.50	0	1.00
15	0.88	0	0.00
16	0.86	0	1.00
17	0.75	0	1.00
18	0.80	0	0.00
19	0.50	0	1.00
20	0.75	0	0.00
21	0.80	0	0.00
22	0.75	0	0.00
23	0.67	0	1.00
24	0.50	0	0.00
25	0.75	0	0.00
Razem	0.61	0.00	0.48

Klasyfikator: MaxEnt kontekst			
Nr. testu	Dokładność	Poprawna	Najlepsza
1	0.80	0	1.00
2	0.80	0	1.00
3	0.80	0	1.00
4	0.50	0	1.00
5	0.80	0	1.00
6	0.67	0	1.00
7	0.80	0	1.00
8	0.67	0	1.00
9	0.67	0	1.00
10	0.83	0	1.00
11	0.80	0	1.00
12	0.67	0	0.00
13	0.67	0	1.00
14	0.50	0	1.00
15	0.88	0	0.00
16	0.86	0	1.00
17	0.75	0	1.00
18	0.80	0	1.00
19	0.50	0	1.00
20	0.75	0	0.00
21	0.80	0	1.00
22	0.75	0	0.00
23	0.67	0	1.00
24	0.50	0	1.00
25	0.75	0	0.00
Razem	0.61	0.00	0.80

Klasyfikator: MaxEnt słownik			
Nr. testu	Dokładność	Poprawna	Najlepsza
1	0.60	0	0.00
2	0.60	1	0.00
3	0.40	1	0.00
4	0.50	1	1.00
5	0.60	0	0.00
6	0.33	1	1.00
7	0.40	0	0.00
8	1.00	1	1.00
9	0.33	0	0.00
10	0.67	0	0.00
11	0.20	1	0.00
12	0.33	0	0.00
13	0.00	0	0.00
14	1.00	1	1.00
15	0.38	1	1.00
16	0.71	1	1.00
17	0.50	0	0.00
18	0.20	1	0.00
19	0.50	0	0.00
20	0.50	1	1.00
21	0.20	0	0.00
22	0.50	0	0.00
23	0.33	1	1.00
24	0.50	1	1.00
25	0.75	0	0.00
Razem	0.38	0.52	0.36

Klasyfikator: MaxEnt razem			
Nr. testu	Dokładność	Poprawna	Najlepsza
1	0.60	0	0.00
2	0.60	1	0.00
3	0.60	1	1.00
4	0.50	1	1.00
5	0.60	0	0.00
6	0.67	1	1.00
7	0.40	0	0.00
8	1.00	1	1.00
9	0.33	0	0.00
10	0.50	0	0.00
11	0.20	1	0.00
12	0.33	0	0.00
13	0.00	0	0.00
14	1.00	1	1.00
15	0.50	1	0.00
16	0.71	1	1.00
17	0.75	0	0.00
18	0.00	0	0.00
19	0.50	0	0.00
20	0.75	1	1.00
21	0.20	0	0.00
22	0.50	0	0.00
23	0.33	1	1.00
24	0.50	1	1.00
25	0.75	0	0.00
Razem	0.41	0.48	0.36

t

Klasyfikator: MaxEnt słowa kluczowe			
Nr. testu	Dokładność	Poprawna	Najlepsza
1	0.80	0	0.00
2	0.80	0	1.00
3	0.80	0	0.00
4	0.50	0	0.00
5	0.80	0	0.00
6	0.67	0	1.00
7	0.80	0	0.00
8	0.67	0	0.00
9	0.67	0	0.00
10	0.83	0	1.00
11	0.80	0	0.00
12	0.67	0	1.00
13	0.67	0	0.00
14	0.50	0	0.00
15	0.88	0	0.00
16	0.86	0	0.00
17	0.75	0	1.00
18	0.80	0	0.00
19	0.50	0	0.00
20	0.75	0	0.00
21	0.80	0	0.00
22	0.75	0	0.00
23	0.67	0	0.00
24	0.50	0	1.00
25	0.75	0	0.00
Razem	0.61	0.00	0.24

Klasyfikator: MaxEnt bez słownika			
Nr. testu	Dokładność	Poprawna	Najlepsza
1	0.80	0	0.00
2	0.80	0	1.00
3	0.80	0	1.00
4	0.50	0	1.00
5	0.80	0	0.00
6	0.67	0	1.00
7	0.80	0	0.00
8	0.67	0	1.00
9	0.67	0	0.00
10	0.83	0	0.00
11	0.80	0	1.00
12	0.67	0	0.00
13	0.67	0	1.00
14	0.50	0	1.00
15	0.88	0	0.00
16	0.86	0	1.00
17	0.75	0	1.00
18	0.80	0	0.00
19	0.50	0	1.00
20	0.75	0	0.00
21	0.80	0	1.00
22	0.75	0	0.00
23	0.67	0	1.00
24	0.50	0	0.00
25	0.75	0	0.00
Razem	0.61	0.00	0.52