

2013 年 CCF 自然语言处理与中文计算会议

中文微博情绪分析评测任务结果

1. 提交结果编号

本次评测共有 19 支队伍提交 58 组有效结果,提交结果编号及所属参评单位对应情况如表 1 所示

编号	所属单位
1	合肥工业大学情感计算与先进智能机器研究院
2	北京语言大学应用语言学研究所
3	华中师范大学中文信息处理实验室
4	中国传媒大学有声媒体语言研究中心
5	重庆理工大学计算机科学与工程学院
6	上海易步信息技术有限公司
7	浙江工商大学计算机与信息工程学院
8	闽江学院
9	山东财经大学计算机科学与技术学院舜耕实验室
10	南京理工大学计算机科学与技术学院
11	中国人民大学信息学院
12	上海交通大学信息内容分析技术国家工程实验室
13	上海交通大学中德语言技术联合实验室
14	清华大学计算机科学与技术系智能技术与系统国家重点实验室
15	武汉大学计算机学院
16	北京工商大学计算机与信息工程学院数据挖掘与社会计算实验室
17	北京信息科技大学智能信息处理研究所
18	大连理工大学软件学院
19	南华大学计算机科学与技术学院

表 1 提交结果编号、系统代号与参评单位对照表

2. 必选任务评测结果

针对每条微博内容,本任务要求判断整条微博是否包含情绪表达(微博情绪判别子任务),对于包含情绪的微博要求参赛队伍判断其整体情绪(情绪分类子任务)。并且根据所使用的资源不同分为资源受限(Close 测试,仅使用组织者提供的词典、分词工具等资源)和资源不受限(Open 测试,可以使用任何的情绪词

典和自然语言处理工具) 两类提交结果。本任务对于情绪句判断任务使用正确率 (Precision), 召回率 (Recall) 和 F 值 (F-measure) 来评价各个参赛队伍提交结果的性能。对于情绪识别任务, 使用宏平均和微平均的准确率、召回率以及 F 值评估各个参赛队伍提交结果的性能。

微博情绪判别子任务 Close 评测和 Open 测试结果如表 2、表 3 所示。

队伍编号	结果序号	正确率	召回率	F 值
1	1	0.4955	0.7318	0.5909
2	1	0.7435	0.453	0.563
3	1	0.7315	0.4542	0.5604
3	2	0.6971	0.4554	0.5509
5	1	0.5257	0.9517	0.6773
5	2	0.5254	0.9507	0.6768
6	1	0.5265	0.9174	0.6691
7	1	0.6451	0.7031	0.6728
9	1	0.3581	0.3322	0.3447
9	2	0.7494	0.6335	0.6866
10	1	0.5338	0.7555	0.6256
11	1	0.5598	0.1033	0.1743
11	2	0.5502	0.1005	0.1699
12	1	0.5325	0.3582	0.4283
13	1	0.5787	0.7257	0.6439
13	2	0.5237	0.9373	0.6719
14	1	0.5747	0.438	0.4971
14	2	0.665	0.71	0.6867
15	1	0.5263	0.9436	0.6757
15	2	0.5263	0.9436	0.6757
16	1	0.6051	0.6741	0.6377
17	1	0.7036	0.3925	0.5039
18	1	0.5226	0.9411	0.672
19	1	0.6676	0.7982	0.7271
19	2	0.6626	0.8012	0.7254

表 2 微博情绪判别子任务 Close 评测结果

队伍编号	结果序号	正确率	召回率	F 值
1	1	0.4955	0.7318	0.5909
4	1	0.6363	0.7616	0.6933
4	2	0.6368	0.7608	0.6933
6	1	0.5335	0.9446	0.6819
7	1	0.6549	0.7044	0.6788

8	1	0.6413	0.8435	0.7286
8	2	0.642	0.8422	0.7286
13	1	0.5808	0.761	0.6588
13	2	0.5245	0.9563	0.6774
14	1	0.665	0.71	0.6867
15	1	0.5213	0.971	0.6784
15	2	0.5213	0.971	0.6784
17	1	0.7053	0.3838	0.4971

表 3 微博情绪判别子任务 Open 评测结果

情绪分类子任务 Close 及 Open 测试的宏平均评测结果及微平均结果如表 4-表 7 所示。

队伍编号	结果序号	正确率	召回率	F 值
1	1	0.2696	0.2186	0.2414
2	1	0.2326	0.124	0.1617
3	1	0.2844	0.169	0.212
3	2	0.2544	0.1481	0.1872
5	1	0.2398	0.2773	0.2572
5	2	0.2296	0.272	0.249
6	1	0.1845	0.2743	0.2206
7	1	0.1722	0.1367	0.1524
9	1	0.1753	0.134	0.1519
9	2	0.2653	0.2201	0.2406
10	1	0.1866	0.1815	0.184
11	1	0.1509	0.0305	0.0507
11	2	0.1583	0.029	0.049
12	1	0.1776	0.0849	0.1149
13	1	0.2572	0.2309	0.2434
13	2	0.1992	0.2719	0.2299
14	1	0.1154	0.0887	0.1003
14	2	0.2474	0.2528	0.2501
15	1	0.2155	0.2803	0.2436
15	2	0.216	0.281	0.2442
16	1	0.2021	0.2151	0.2084
17	1	0.2503	0.131	0.172
18	1	0.1959	0.276	0.2291
19	1	0.2704	0.3064	0.2873
19	2	0.2109	0.1996	0.2051

表 4 情绪分类子任务 Close 宏平均评测结果

队伍编号	结果序号	正确率	召回率	F 值
1	1	0.2696	0.2186	0.2414
4	1	0.2842	0.348	0.3129
4	2	0.2842	0.3469	0.3124
6	1	0.1927	0.3037	0.2358
7	1	0.2245	0.1357	0.1692
8	1	0.2251	0.2977	0.2564
8	2	0.2288	0.2996	0.2595
13	1	0.2732	0.2639	0.2684
13	2	0.2252	0.2946	0.2553
14	1	0.2474	0.2528	0.2501
15	1	0.2093	0.2836	0.2409
15	2	0.2122	0.2853	0.2434
17	1	0.2421	0.1272	0.1668

表 5 情绪分类子任务 Open 宏平均评测结果

队伍编号	结果序号	正确率	召回率	F 值
1	1	0.2263	0.3342	0.2698
2	1	0.2862	0.1744	0.2167
3	1	0.3834	0.238	0.2937
3	2	0.3122	0.2039	0.2467
5	1	0.2196	0.3976	0.283
5	2	0.2112	0.3821	0.272
6	1	0.2046	0.3565	0.26
7	1	0.2188	0.2384	0.2282
9	1	0.1907	0.1769	0.1836
9	2	0.3655	0.309	0.3349
10	1	0.2135	0.3021	0.2501
11	1	0.2692	0.0497	0.0838
11	2	0.2578	0.0471	0.0796
12	1	0.2114	0.1255	0.1575
13	1	0.2667	0.3344	0.2967
13	2	0.2114	0.3783	0.2712
14	1	0.1457	0.1109	0.126
14	2	0.2924	0.3121	0.3019
15	1	0.2164	0.388	0.2778
15	2	0.2148	0.3852	0.2758
16	1	0.2112	0.2353	0.2226
17	1	0.2914	0.1626	0.2087
18	1	0.215	0.3872	0.2765
19	1	0.3133	0.3746	0.3412
19	2	0.2539	0.307	0.2779

表 6 情绪分类子任务 Close 微平均评测结果

队伍编号	结果序号	正确率	召回率	F 值
1	1	0.2263	0.3342	0.2698
4	1	0.3232	0.3868	0.3521
4	2	0.3231	0.386	0.3518
6	1	0.2153	0.3813	0.2752
7	1	0.2228	0.2396	0.2309
8	1	0.2765	0.3637	0.3142
8	2	0.2781	0.3647	0.3155
13	1	0.2738	0.3588	0.3106
13	2	0.2145	0.3911	0.2771
14	1	0.2924	0.3121	0.3019
15	1	0.2125	0.3959	0.2766
15	2	0.2108	0.3927	0.2744
17	1	0.2848	0.1551	0.2008

表 7 情绪分类子任务 Open 微平均评测结果

3. 可选任务评测结果

针对微博的各个句子，可选任务要求参赛队伍识别出句子的主要情绪和次要情绪，同样根据资源受限情况分为 Close 和 Open 两类测试。此任务使用平均精度作为评价指标，在该指标下，对于参赛队伍提交结果，若某情绪排序位置在答案位置之前则认为该情绪识别正确。在宽松指标下，主要情绪和次要情绪识别正确均可加 1 分，在严格指标下，正确识别主要情绪加 1 分，次要情绪加 0.5 分。对于所有情绪识别结果求算术平均值获得最终的平均精度。可选任务 Close 类评测结果如表 8 所示：

队伍编号	结果序号	平均精度（宽松指标）	平均精度（严格指标）
5	1	0.1486	0.1459
5	2	0.1494	0.1467
6	1	0.1691	0.1622
12	1	0.0233	0.0225
14	1	0.0605	0.0586
15	1	0.1883	0.1806
15	2	0.1827	0.1792
16	1	0.1648	0.1585
16	2	0.1648	0.1585
17	1	0.0284	0.0275
17	2	0.2054	0.1965
18	1	0.1848	0.1769

表 8 可选任务 Close 评测结果

Open 类评测结果如表 9 所示:

队伍编号	结果序号	平均精度 (宽松指标)	平均精度 (严格指标)
4	1	0.1827	0.1769
4	2	0.1815	0.1757
6	1	0.1861	0.1784
8	1	0.178	0.1749
8	2	0.1831	0.1799
15	1	0.2077	0.1952
17	1	0.0391	0.0379
17	2	0.1486	0.1415

表 9 可选任务 Open 评测结果