

## 2012 年 CCF 自然语言处理与中文计算会议

### 中文微博情感分析评测结果

#### 1. 提交结果编号

本次评测共有 34 支队伍提交 53 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。

表 1 提交结果编号与参评单位对照表

提交结果编号	参评单位
1	北京工商大学
2	北京工商大学
3	北京航空航天大学计算机学院
4	北京航空航天大学计算机学院
5	北京理工大学海量语言信息处理与云计算应用工程技术研究中心 <sup>1</sup>
6	北京理工大学网络搜索挖掘与安全实验室
7	北京理工大学海量语言信息处理与云计算应用工程技术研究中心 <sup>2</sup>
8	北京理工大学海量语言信息处理与云计算应用工程技术研究中心 <sup>2</sup>
9	大连理工大学
10	大连理工大学
11	广东工业大学 DMIR 实验室
12	哈尔滨工业大学语言技术研究中心网络智能研究室
13	哈尔滨工业大学语言技术研究中心网络智能研究室
14	哈尔滨工业大学计算机科学与技术学院/机器智能与翻译研究室
15	哈尔滨工业大学计算机科学与技术学院/机器智能与翻译研究室
16	哈尔滨工业大学（威海）
17	海军工程大学信息安全系
18	黑龙江大学计算机科学技术学院
19	湖南工业大学计算机与通信学院
20	湖南工业大学计算机与通信学院
21	湖南科技大学外国语学院
22	华侨大学计算机科学与技术学院
23	华侨大学计算机科学与技术学院
24	华中科技大学
25	南京大学计算机科学与技术系自然语言处理研究组
26	南京理工大学
27	南京理工大学
28	清华大学计算机系智能技术与系统国家重点实验室信息检索组
29	清华大学计算机系智能技术与系统国家重点实验室信息检索组

<sup>1</sup>参评队伍联系人为刘全超

<sup>2</sup>参评队伍联系人为王金刚

提交结果编号	参评单位
30	厦门大学人工智能研究所
31	厦门大学人工智能研究所
32	上海交通大学中德语言技术联合实验室
33	上海交通大学中德语言技术联合实验室
34	上海交通大学计算机系
35	上海交通大学计算机系
36	上海交通大学信息内容分析技术国家工程实验室
37	上海交通大学信息内容分析技术国家工程实验室
38	同济大学计算机科学与技术系
39	武汉大学计算机学院
40	武汉大学计算机学院
41	西安交通大学/山西省天地网技术重点实验室
42	西南大学计算机信息科学学院 1010 实验室
43	浙江大学城市学院
44	浙江工商大学计算机与信息工程学院
45	浙江工商大学计算机与信息工程学院
46	郑州大学自然语言处理实验室
47	郑州大学自然语言处理实验室
48	中国传媒大学国家语言资源监测与研究中心有声媒体语言分中心
49	中国传媒大学国家语言资源监测与研究中心有声媒体语言分中心
50	中国科学院声学研究所
51	中国科技大学知识与数据工程实验室
52	中国科技大学知识与数据工程实验室
53	北京交通大学自然语言处理研究室

## 2. 观点句识别评测结果

针对每条微博中的各个句子，本任务要求判断出该句是观点句还是非观点句。评测使用正确率（Precision），召回率（Recall）和 F 值（F-measure）来评价各个参赛队伍对观点句的识别结果。微平均以整个数据集为一个评价单元，计算整体的评价指标；宏平均以每个话题为一个评价单元，计算参评系统在该话题中的评价指标，最后计算所有话题上各指标的平均值。评测结果如表 2 所示。

表 2 观点句识别评测结果

结果编号	微平均			宏平均		
	正确率	召回率	F 值	正确率	召回率	F 值
1	0.743	0.691	0.716	0.745	0.680	0.707
2	0.733	0.746	0.739	0.734	0.735	0.731
3	0.681	0.918	0.782	0.684	0.914	0.779
4	0.707	0.577	0.635	0.715	0.577	0.632
5	0.673	0.773	0.720	0.680	0.774	0.712
6	0.742	0.694	0.717	0.706	0.684	0.692
7	0.740	0.557	0.636	0.733	0.541	0.618

结果编号	微平均			宏平均		
	正确率	召回率	F 值	正确率	召回率	F 值
8	0.734	0.528	0.614	0.724	0.514	0.599
9	0.825	0.603	0.697	0.828	0.589	0.679
10	0.822	0.592	0.688	0.824	0.581	0.674
11	0.835	0.449	0.584	0.836	0.435	0.557
12	0.738	0.726	0.732	0.743	0.717	0.726
13	0.738	0.726	0.732	0.743	0.717	0.726
14	0.619	0.378	0.469	0.601	0.383	0.413
15	0.619	0.378	0.469	0.601	0.383	0.413
16	0.647	0.757	0.697	0.648	0.751	0.689
17	0.828	0.537	0.651	0.826	0.520	0.629
18	0.728	0.502	0.594	0.736	0.502	0.582
19	0.747	0.439	0.553	0.741	0.431	0.542
20	0.781	0.406	0.534	0.776	0.395	0.520
21	0.746	0.772	0.759	0.747	0.757	0.748
22	0.707	0.656	0.681	0.713	0.651	0.672
23	0.700	0.733	0.716	0.705	0.734	0.713
24	0.737	0.536	0.621	0.743	0.522	0.607
25	0.695	0.473	0.563	0.695	0.461	0.548
26	0.745	0.406	0.525	0.742	0.394	0.503
27	0.745	0.406	0.525	0.742	0.394	0.503
28	0.714	0.717	0.716	0.722	0.708	0.704
29	0.715	0.745	0.729	0.721	0.738	0.721
30	0.740	0.646	0.690	0.744	0.639	0.680
31	0.733	0.683	0.707	0.737	0.678	0.702
32	0.671	0.944	0.784	0.674	0.942	0.783
33	0.671	0.944	0.784	0.674	0.942	0.783
34	0.805	0.588	0.680	0.807	0.581	0.671
35	0.745	0.789	0.767	0.748	0.782	0.760
36	0.674	0.891	0.768	0.679	0.892	0.764
37	0.660	0.871	0.751	0.663	0.869	0.747
38	0.704	0.562	0.625	0.699	0.557	0.615
39	0.725	0.632	0.675	0.723	0.618	0.661
40	0.708	0.649	0.677	0.708	0.634	0.663
41	0.638	0.221	0.328	0.630	0.217	0.320
42	0.783	0.338	0.472	0.792	0.337	0.452
43	0.780	0.455	0.575	0.781	0.443	0.557
44	0.696	0.348	0.464	0.686	0.348	0.446
45	0.645	0.959	0.772	0.649	0.960	0.770
46	0.765	0.647	0.701	0.760	0.641	0.680
47	0.779	0.542	0.639	0.767	0.529	0.615
48	0.756	0.802	0.779	0.758	0.788	0.769

结果编号	微平均			宏平均		
	正确率	召回率	F 值	正确率	召回率	F 值
49	0.756	0.812	0.783	0.757	0.797	0.773
50	0.773	0.119	0.206	0.766	0.112	0.181
51	0.728	0.658	0.691	0.732	0.651	0.686
52	0.716	0.716	0.716	0.719	0.712	0.711
53	0.701	0.334	0.452	0.707	0.341	0.454

### 3. 情感倾向性判断评测结果

本任务要求判断微博中每条观点句的情感倾向。本任务同样使用正确率（Precision），召回率（Recall）和 F 值（F-measure）作为评价标准。评测结果如表 3 所示。

表 3 情感倾向性判断评测结果

结果编号	微平均			宏平均		
	正确率	召回率	F 值	正确率	召回率	F 值
1	0.831	0.574	0.679	0.823	0.563	0.666
2	0.824	0.614	0.704	0.825	0.608	0.698
3	0.761	0.698	0.728	0.768	0.702	0.733
4	0.764	0.440	0.559	0.758	0.445	0.559
5	0.734	0.568	0.640	0.738	0.574	0.642
6	0.782	0.565	0.656	0.783	0.562	0.653
7	0.724	0.403	0.518	0.708	0.387	0.496
8	0.718	0.379	0.496	0.703	0.365	0.477
9	0.841	0.507	0.633	0.849	0.497	0.620
10	0.833	0.493	0.619	0.843	0.487	0.611
11	0.426	0.426	0.426	0.413	0.413	0.413
12	0.881	0.640	0.741	0.878	0.632	0.733
13	0.863	0.626	0.726	0.860	0.619	0.718
14	0.258	0.097	0.141	0.341	0.105	0.139
15	0.261	0.099	0.143	0.342	0.107	0.141
16	0.559	0.559	0.559	0.561	0.561	0.561
17	0.772	0.415	0.540	0.776	0.404	0.523
18	0.809	0.406	0.541	0.791	0.407	0.530
19	0.598	0.262	0.365	0.583	0.253	0.350
20	0.594	0.241	0.343	0.578	0.232	0.328
21	0.796	0.614	0.693	0.789	0.600	0.679
24	0.643	0.344	0.449	0.641	0.335	0.437
25	0.803	0.379	0.515	0.800	0.370	0.502
26	0.647	0.399	0.493	0.641	0.390	0.482
28	0.788	0.565	0.658	0.780	0.562	0.649
29	0.794	0.591	0.678	0.786	0.590	0.671
30	0.740	0.478	0.580	0.734	0.472	0.572
31	0.725	0.495	0.588	0.725	0.490	0.583

结果编号	微平均			宏平均		
	正确率	召回率	F 值	正确率	召回率	F 值
34	0.893	0.481	0.625	0.895	0.481	0.622
35	0.886	0.631	0.737	0.888	0.630	0.733
36	0.587	0.587	0.587	0.579	0.579	0.579
37	0.850	0.850	0.850	0.854	0.854	0.854
38	0.691	0.389	0.498	0.693	0.387	0.491
39	0.809	0.511	0.627	0.799	0.496	0.607
40	0.740	0.480	0.582	0.731	0.465	0.565
41	0.832	0.184	0.301	0.829	0.181	0.293
42	0.288	0.288	0.288	0.289	0.289	0.289
43	0.879	0.400	0.550	0.872	0.391	0.533
44	0.803	0.266	0.399	0.733	0.263	0.375
45	0.804	0.771	0.787	0.809	0.778	0.793
46	0.902	0.584	0.709	0.899	0.578	0.690
47	0.857	0.464	0.602	0.855	0.452	0.579
48	0.842	0.675	0.749	0.840	0.663	0.739
49	0.844	0.685	0.756	0.842	0.672	0.745
50	0.108	0.108	0.108	0.102	0.102	0.102
51	0.476	0.341	0.397	0.459	0.329	0.382
52	0.476	0.341	0.397	0.459	0.329	0.382
53	0.450	0.150	0.225	0.435	0.151	0.222

#### 4. 情感要素抽取评测结果

本任务要求找出微博中每条观点句作者的评价对象，即情感对象，同时判断针对情感对象的观点极性。本任务采用严格评价和宽松评价两种方式，均使用准确率、召回率以及 F 值作为评价标准。

##### 4.1 严格评价指标

在严格评价中，要求提交的情感对象的 offset 和答案完全相同并且情感对象极性也相同时才算正确。评测结果如表 4 所示。

表 4 情感要素抽取（严格评价指标）评测结果

结果编号	微平均			宏平均		
	正确率	召回率	F 值	正确率	召回率	F 值
3	0.066	0.147	0.091	0.070	0.144	0.093
4	0.065	0.102	0.080	0.068	0.101	0.080
6	0.028	0.039	0.032	0.031	0.040	0.035
9	0.485	0.066	0.116	0.474	0.066	0.113
10	0.425	0.077	0.131	0.417	0.076	0.126
17	0.311	0.177	0.225	0.324	0.174	0.220
21	0.081	0.005	0.010	0.031	0.005	0.009
24	0.120	0.089	0.102	0.133	0.090	0.105

结果编号	微平均			宏平均		
	正确率	召回率	F 值	正确率	召回率	F 值
25	0.182	0.100	0.129	0.196	0.101	0.132
32	0.160	0.160	0.160	0.172	0.169	0.170
33	0.112	0.112	0.112	0.121	0.118	0.119
38	0.006	0.003	0.004	0.006	0.003	0.004
39	0.135	0.078	0.099	0.149	0.081	0.103
40	0.120	0.071	0.089	0.125	0.073	0.091
44	0.126	0.027	0.045	0.121	0.027	0.042
45	0.106	0.062	0.078	0.109	0.064	0.081
46	0.260	0.164	0.201	0.271	0.172	0.205
47	0.117	0.058	0.077	0.123	0.062	0.081
48	0.303	0.273	0.288	0.306	0.263	0.278
49	0.303	0.275	0.288	0.305	0.265	0.278
50	0.241	0.063	0.100	0.244	0.058	0.084
53	0.031	0.009	0.014	0.029	0.010	0.015

#### 4.2 宽松评价指标

在宽松评价中，评价指标通过提交的结果与答案之间的覆盖率计算。评测结果如表 5 所示。

表 5 情感要素抽取（宽松评价指标）评测结果

结果编号	微平均			宏平均		
	正确率	召回率	F 值	正确率	召回率	F 值
3	0.264	0.325	0.291	0.269	0.330	0.291
4	0.241	0.213	0.226	0.246	0.219	0.227
6	0.086	0.073	0.078	0.092	0.074	0.080
9	0.636	0.086	0.152	0.643	0.086	0.149
10	0.569	0.099	0.169	0.572	0.098	0.165
17	0.404	0.223	0.287	0.420	0.219	0.279
21	0.137	0.007	0.013	0.071	0.007	0.012
24	0.198	0.138	0.163	0.219	0.139	0.167
25	0.279	0.140	0.186	0.286	0.138	0.183
32	0.290	0.220	0.250	0.302	0.229	0.257
33	0.287	0.189	0.228	0.300	0.196	0.234
38	0.107	0.028	0.045	0.110	0.029	0.045
39	0.232	0.105	0.145	0.243	0.106	0.145
40	0.254	0.114	0.157	0.252	0.113	0.153
44	0.257	0.040	0.070	0.239	0.040	0.066
45	0.236	0.098	0.139	0.241	0.102	0.142
46	0.398	0.249	0.307	0.400	0.258	0.301
47	0.201	0.110	0.142	0.201	0.115	0.136
48	0.388	0.354	0.371	0.393	0.342	0.359

结果编号	微平均			宏平均		
	正确率	召回率	F 值	正确率	召回率	F 值
49	0.387	0.356	0.371	0.391	0.344	0.359
50	0.422	0.082	0.137	0.387	0.075	0.115
53	0.091	0.019	0.032	0.093	0.020	0.032