中文微博情感分析评测大纲

1. 评测对象

本次评测的对象是面向中文微博的情感分析中的核心技术，包括观点句识别、情感倾向性分析和情感要素抽取。

2. 任务设置

本评测设置了如下 3 个子任务，其中任务 1 是必选任务，任务 2 和任务 3 都是基于任务 1 的，参赛队伍可以选作。

2.1 观点句识别

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。

提交格式：

id　　run-tag　　weibo-id　　sentence-id　　opinion

说明

id：结果序号

run-tag：队伍标识

weibo-id：微博 id

sentence-id：句子 id

opinion：观点句标识，是观点句则为 Y，非观点句则为 N

例如，如下两条微博：

weibo1：

```
<weibo id="1">
  <sentence id="1" >渭南城管撕春联事件在成都公交车上的分众传媒大量报道！
</sentence >
  <sentence id="2"> 渭南城管伤不起啊！</sentence >
</weibo>
```

weibo2:

```
<weibo id="2">
  <sentence id="1">LinkedIn 在 2011 年 6 月份的时候，中国大陆地区用户大约有
100 万，目前则是 200 多万。</sentence>
  <sentence id="2">LinkedIn 应该没做过一分钱的广告吧，或许 IPO 就是非常好的
宣传广告。</sentence>
</weibo>
```

weibo1 中有两个句子，第一句是非观点句，第二句是观点句。weibo2 中有 4 个句子，其中第一句是非观点句，第二句观点句。则正确的输出结果为：

1 xyz 1 1 N

2 xyz 1 2 Y

3 xyz 2 1 N

4 xyz 2 2 Y

注：本评测的观点句的定义不包括表达自我情感的句子，比如“我感到很高兴。”这样的句子是情感句，但不属于本评测定义的观点句。本评测定义的观点句只限定于对其它对象的评价（例如“我真心喜欢 iphone 的屏幕效果。”），不包括内心自我情感。

不同字段之间用\t 隔开，下同。

评价标准：

本任务使用正确率（Precision），召回率（Recall）和 F 值（F-measure）来评价各个参赛队伍的系统。其计算公式如下：

$$Precision = \frac{\#system\_correct}{\#system\_proposed}$$

$$\mathrm{Recall} = \frac{\#system\_correct}{\#\mathrm{gold}}$$

$$F-measure = \frac{2\times \mathrm{Precision}\times \mathrm{Recall}}{\mathrm{Precision}+\mathrm{Recall}}$$

#gold 是人工标注的结果

2.2 情感倾向性判断

本任务要求判断微博中每条观点句的情感倾向。观点句的情感倾向可以分为正面（positive），负面（negative）和中性（neural）。

提交格式：

id run-tag weibo-id sentence-id polarity

说明

id：结果序号

run-tag：队伍标识

weibo-id：微博 id

sentence-id：观点句 id

polarity：情感倾向标识，正面为 POS，负面为 NEG，中性为 NEU。

比如：上面 weibo1 和 weibo2 两条微博中，weibo1 的第二句是观点句，情感倾向为中性。weibo2 的第二句是观点句，情感倾向为正面。则其结果应如下：

1 xyz 1 2 NEU

2 xyz 2 2 POS

注：对于无法明确分类到正面或负面的观点句应归类到中性类别中。

评价标准：

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。

$$Precision = \frac{\#system\_correct(polarity = POS, NEG, NEU)}{\#system\_proposed(opinion = Y)}$$

$$\mathrm{Recall} = \frac{\#system\_correct(polarity = POS, NEG, NEU)}{\#\mathrm{gold(opinion=Y)}}$$

$$F - measure = \frac{2 \times \mathrm{Precision} \times \mathrm{Recall}}{\mathrm{Precision} + \mathrm{Recall}}$$

2.3 情感要素抽取

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。每条观点句可能含有若干情感对象。同时判断针对情感对象的观点极性。

注：

1. 只对微博中的观点句进行情感对象的抽取。
2. 本任务属于句子级任务，情感对象只能从当前句子中抽取。对于那些没有情感对象(有些情况情感对象是隐含的)的观点句，参赛队伍可以不进行抽取。
3. 对于情感对象文本长短的定义本评测采用灵活的方式，比如“渭南城管伤不起啊!”情感对象可以是：“城管”或“渭南城管”；但“南城”或“南城管”则不属于本例句的情感对象。
4. 对于情感对象是代词的情况，参赛队伍可以做指代消解，也可以不做。比如“小明就读于北京大学，他是名优秀的学生。”情感对象可以是“他”或“小明”

提交格式

id run-tag weibo-id sentence-id begin-offset end-offset polarity 说明

id：结果序号

run-tag：队伍标识

weibo-id：微博 id

sentence-id：句子 id

begin-offset: 情感对象的起始位置

end-offset：情感对象的终止位置

polarity：对情感对象的观点极性，POS 代表正面，NEG 代表负面， NEU 代表中性。

比如 weibo1 和 weibo2 的情感要素如下：

1	xyz	1	2	0	3	NEU
2	xyz	2	2	0	7	NEG
3	xyz	2	2	24	16	POS

注：对于无法明确分类为正面或负面的观点极性应归类为中性。文件采用 unicode（utf-16）编码，每个字符都占两个字节，任意句子中第一个字符的 offset 为 0，第二个字符的 offset 为 1，以此类推。比如：weibo1 第二句开始位置的“渭南城管”这四个字符对应的 offset 分别为 0,1,2,3。

评价标准：

本任务同样采用准确率（Precision）、召回率（Recall）以及 F 值（F-measure）作为评价标准。

3. 评测数据集

本次评测数据来自腾讯微博1。评测数据全集包括 20 个话题，每个话题采集大约 1000 条微博，共约 20000 条微博。数据采用 xml 格式，已经预先切分好句子。

样例：

weibo1：

```
<weibo id="1">
  <sentence id="1" >渭南城管撕春联事件在成都公交车上的分众传媒大肆报道！
</sentence >
  <sentence id="2"> 渭南城管伤不起啊！ </sentence >
</weibo>
```

weibo2:

```
<weibo id="2">
  <sentence id="1">LinkedIn 在 2011 年 6 月份的时候，中国大陆地区用户大约有
100 万，目前则是 200 多万。</sentence>
  <sentence id="2">LinkedIn 应该没做过一分钱的广告吧，或许 IPO 就是非常好的
宣传广告。</sentence>
```

[1] http://t.qq.com/

</weibo>

其中，每条微博对应一个*<weibo>*元素，每个句子对应一个*<sentence>*元素。文件采用unicode（utf-16）编码。

4. 评测方法

本次评测为离线评测。参评单位自行处理数据，生成相应结果后提交。答案采用人工标注的方法确定。参赛单位需要处理全部评测数据，但用于实际评测的人工标注数据仅为评测数据全集的 10%左右。

具体评测步骤为：

1) 评测单位预先提供测试样例（包括答案）
2) 评测单位给出测试数据
3) 参评单位运行被测系统，得出测试结果
4) 参评单位提交测试结果
5) 评测单位标注答案，运行自动评测程序，统计评测结果

5. 评测要求

参评单位应当采用自动的方法，针对微博进行情感分析。参评系统应当预先训练模型、调整好所有参数，运行过程中不得有人工干预。本次评测不限制使用各种语义资源。对于每个子任务，参评单位至多提交 2 组结果。

6. 评测日程

2012/1/1-2/29：起草评测大纲，征求各方意见；

2012/3/1-3/31：修订完善评测大纲，确定评测数据；

2012/4/1-30：发布评测任务，接受评测报名；

2012/5/1：发布评测样例数据集；

2012/5/2-7/15：构建评测数据集，制定标准答案；

2012/7/16-7/31：发布评测数据集，提交运行结果；

2012/8/1-8/31：组织专家评测小组进行结果评判，发布评测结果；

2012/9/1-9/24：征集评测论文；

2012/9/25-9/30：确定受邀报告；

2012/10/29-11/6：宣读报告，交流经验和技术。

7. 如何注册

参加评测的单位需要在接受报名的时间内到如下评测主页下载报名表，并在截止日期前通过电子邮件或传真方式发送给评测组织者。报名应以研究机构或公司为单位，暂不接收个

人报名。

http://tcci.ccf.org.cn/conference/2012/

如果你有任何关于本次评测的问题请发邮件至：huangxiaojiang@pku.edu.cn

8. 本次评测的组织

- **主办单位**

  中国计算机学会中文信息技术专业委员会（CCF TCCI）

- **承办单位**

  北京大学，MSRA

- **协办单位**

  数字出版国家重点实验室

- **评测委员会（按照姓氏拼音排序）**

  李寿山、刘群、万小军、韦福如、徐睿峰、吴云芳