

跨语言情感分类评测大纲

1. 引言

情感分类随着网络评论的海量增长而迅速兴起,其研究价值和应用价值受到人们越来越多的重视。情感分类系统通常依赖于标注语料并结合分类算法来实现。然而,情感标注语料的分布在不同语言下是极不均衡的。因此在当前语言的标注语料缺乏时,利用其他语言的资源来实现情感分类已经成为了一个热门的研究课题。本任务的评测内容既为通过英语的情感资源实现中文评论的情感分类。

2. 任务介绍

本任务要求参赛队伍在仅利用组织方提供的英文资源的前提下,对测试集内的每条中文评论进行倾向性分类。

参赛队伍需针对测试集提交统一的结果文件。文件内每行包含一条评论的情感分类结果,其格式为:

id team-tag review-id polarity

id: 提交结果序号

run-tag: 参赛队伍标签

review-id: 中文评论标号

polarity: 分类结果。如果是正向则为“P”, 否则为“N”

注: run-tag 的格式为“队伍标识_提交结果组号”, 队伍标识可自定, 组号用于区分同一队伍的多组提交结果。不同字段之间用\t 隔开。

3. 评价标准

本任务使用准确率(accuracy)作为评价指标。

$$Accuracy = \frac{\#system_correct}{\#system_total}$$

其中 $\#system_correct$ 表示分类准确的评论数, $\#system_total$ 表示测试集内全部评论数。

4. 评测数据

本任务的所有数据均由主办方提供，主要包含三部分：1) 英文标注数据和英文情感词典；2) 中文未标注语料；3) 中文测试集。数据均采用 XML 格式，UTF-8 编码存储。

4.1 英文标注数据

英文标注数据为 Amazon.com 的用户评论，共包含 3 个不同领域：DVD，书籍，音乐。每个领域内有 2000 条正面评价和 2000 条负面评价。

数据文件的组织结构为 XML 格式，每条评论对应一个<item>元素。一条评论内包含 4 个子元素：评论摘要(<summary>)、评论内容(<text>)、评论倾向(<polarity>)和评论类别(<category>)。图 1 列举了一条书籍的正向评论。

```
<item>
  <summary>Just found my childhood copy.</summary>
  <text>I loved this book as a child. I read it over and over. The pictures
were great. Sort of scary but that was half the fun</text>
  <polarity>P</polarity>
  <category>books</category>
</item>
```

图 1 英文标注语料

4.2 少量中文标注数据

提供包含 DVD，书籍，音乐三个领域的少量中文标注数据，正负例各 20 条。

4.3 英文情感词典

提供 MPQA 情感词典(Multi-Perspective Question Answering Subjectivity Lexicon)¹，共包含 2789 个正向词和 6079 个负向词。同时 MPQA 情感词典还给出了每个词的情感强度，词性等信息。

4.4 中文未标注语料

针对 DVD，书籍和音乐每个领域提供大量中文未标注评论。未标注语料的数据格式与英文标注数据类似，但每条评论下无<polarity>元素。

4.5 中文测试集

¹ http://mpqa.cs.pitt.edu/#subj_lexicon

DVD，书籍和音乐每个领域均包含 4000 条测试评论，其中正负向评论均为 2000 条。测试集同样采用 XML 格式，每条评论对应一个<item>元素。一条评论内包含 4 个子元素：评论摘要(<summary>)、评论内容(<text>)、评论标号(<review_id>)和评论类别(<category>)。

5. 评测要求

1) 除组织方提供的数据外，不得使用任何中英文标注数据、中英文未标注数据和中英文情感词典等资源。

2) 不得人工干预评测结果，不得对测试集进行人工标注。

6. 评测方法

本次评测为离线评测。参评单位自行处理数据，按照规定格式生成相应结果后提交。

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；

7. 评测日程

2013/2/28：发布评测大纲；

2013/3/31：评测报名截止；

2013/4/30：发布样例数据；

2013/5/25：发布测试数据；

2013/5/31：参赛队伍提交运行结果；

2013/6/15：发布评测结果、征集学术论文；

2013/7/31：提交学术论文；

2013/8/31：发布学术论文录用通知、征集评测报告；

2013/9/15: 提交学术论文终稿;

2013/9/30: 确定评测报告;

2013/11: 参加会议并宣读学术论文与评测报告;

8. 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。

<http://tcci.ccf.org.cn/conference/2013/>

报名应以研究机构或公司为单位,暂不接收个人报名。如果你有任何关于本次评测的问题请发邮件至周新杰(zhouxinjie@pku.edu.cn)。

9. 评测委员会 (按照姓氏拼音排序)

冯岩松、李寿山、李智星、万小军、徐睿峰、张梅山