

第八届中文倾向性分析评测（COAE2016）大纲

信息检索专委会评测委员会

1. 引言

文本倾向性（观点和情感等）分析已连续多年成为自然语言处理领域研究的热点问题之一。在 SIGIR、ACL、WWW、CIKM、WSDM 等著名国际会议上，针对这一问题的研究成果层出不穷。随着研究的深入展开，也出现了一些新的研究关注点，如 Aspect-Based Opinion Mining, Context-sensitive Opinion Mining 等。在国内，对于文本倾向性分析的研究正处于快速发展中。前七届中文倾向性分析评测（COAE2008-2015）持续在词语级、句子级、篇章级进行了中文倾向性分析评测，任务涉及主客观分析、情感极性分析、评价对象抽取以及搭配抽取、比较句和否定句抽取等方面。通过开展评测，推动和加速了中文倾向性分析研究的发展。如何结合中文处理的特点，进一步推动中文倾向性分析的发展仍是需要深入思考的问题。因此，在前七届中文倾向性分析评测的基础上，本届评测将继续致力于探索中文倾向性分析的新技术、新方法，开展微博观点摘要、用户评论分析等任务，以推动中文倾向性分析理论和技术的研究与应用，在此基础上，建立并完善中文倾向性分析研究的基础数据集和评测标准，开展评测资源共享服务。

此外，知识图谱近年来成为新的研究热点，关于知识建模、知识抽取、知识图谱构建、知识图谱计算的研究工作大量出现。倾向性分

析的评价对象抽取和搭配抽取任务与知识图谱中的知识抽取工作有很大的相关性，为此，我们在今年新增了面向知识抽取的关系分类和关系对抽取评测任务，既是对 CCIR 评测的扩展，也是对中文信息检索技术的进一步推动。

今年的评测任务不限定使用外部资源，但需要明确指出使用的情感词典、情感本体库等资源，以便能够考察评测模型与方法本身的有效性。同时，评测技术论文需按照 CCIR2016 的要求与格式来撰写，投稿论文将按照 CCIR2016 的大会论文标准进行专家评审，享受同其它投稿论文一样的推荐和录用机会。

2. 评测任务设置

第八届 COAE2016 评测在前七届中文倾向性评测分析的基础上重点对于微博观点摘要、用户评论的倾向与方面进行评测，同时增设关系分类和关系对抽取评测任务。

具体评测任务设置如下表所示：

任务号	类型	任务名称	任务描述
任务 1	篇章级	微博观点摘要	给定若干话题，每个话题给定一个内容相关的微博文档集，生成话题相关的观点文摘
任务 2	篇章级 - 句子级 - 词语级	用户评论的倾向与方面	给定若干影视评论，完成子任务 1-3： 1. 判断每个影视评论的总体倾向（正面、负面、中立） 2. 判断影视评论中每个短句所属的方面（剧情、导演、演员、人物、配乐、画面、其他） 3. 判断影视评论中每个短句针对所属方面的倾向性（正面、负面、中立），抽取倾向性判断的描述词语
任务 3	句子级-	关系分类和	给定若干句子，完成子任务 1-2：

	实体级	实体对抽取	1. 识别出包含实体关系的句子并进行实体关系类型分类。 2. 抽取每个包含实体关系的句子中具有特定关系的实体对。
--	-----	-------	---

各评测任务详细说明如下。

2.1 微博观点摘要

微博是网民表达观点的重要平台，也是了解网民观点的重要渠道。如何基于倾向性分析技术和多文档摘要技术，综合构建话题的微博观点文摘是现阶段文本倾向性分析中最具挑战的问题之一，具有很好的研究意义和应用价值。

本项任务是给定若干个话题以及每个话题相关的一个微博多文档集，要求参赛系统自动分析文档集中与话题相关的观点句子，构建**不超过 10 句**的微博观点文摘。微博文档集是指来自微博的若干篇博文组成的文本集合。该任务是倾向性分析与多文档摘要的组合任务。

【提交结果格式】

Id Run-tag TopicID Digest

Id: 任务编号（此任务编号为 1）

Run-tag: 评测参与者编号

TopicID: 话题编号

Digest: 观点摘要

注意：不同字段以 tab 分割

【提交文档格式】

<TopicID> XXXX </TopicID>

< DocID> ZZZZ</ DocID>

<Document> DDDDDDDDDD </ Document>

其中，

TopicID: 话题编号

DocID: 文档集编号

Document: 文档内容

【任务示例】

<TopicID>0005 </TopicID>

<DocID> 0001</ DocID>

<Document>

<Doc1-1>电池会爆炸的。</Doc1-1>

<Doc1-2>散热电池气炸你。</Doc1-2>

<Doc1-3>三星是血泪史，无法找到一个准确的词来形容它的卡。

</Doc1-3>

<Doc1-4> 建议不要，很卡，真心的[微笑]。</Doc1-4>

<Doc1-5> 三星真是越做越丑了。</Doc1-5>

<Doc1-6> 三星大边框傻吧。</Doc1-6>

<Doc1-7> 丑，韩国棒子觉得随便做的产品在中国赚钱，垃圾产品。</Doc1-7>

<Doc1-8>我觉得 note7 唯一的遗憾就是国行不提供 google service, 不然就圆满了日。</Doc1-8>

<Doc1-9>真心不推荐。</Doc1-9>

<Doc1-10>跟 MIUI 相比差的真的不是一点半点[拜拜]。
</Doc1-10>

<Doc1-11>密恐已死，三星你看着办吧。</Doc1-11>

<Doc1-12>垃圾，难看的影响市容，都拿不出手。</Doc1-12>

<Doc1-13>丑是一定的。</Doc1-13>

<Doc1-14>眼晕。</Doc1-14>

<Doc1-15> miui? 那图标好看吗? 国产也不是那么支持的吧?
米 UI 好用是真的。</Doc1-15>

<Doc1-16>...</Doc1-16>

</ Document>

注释：编号 0005 为话题编号，对应话题为“三星 Note7”

结果输出：

1 ***** 0005 垃圾，难看的影响市容，都拿不出手。真心不推荐。三星大边框傻吧。丑是一定的。跟 MIUI 相比差的真的不是一点半点。建议不要，很卡，真心的。

其中，

Id: 1

Run-tag: *****

TopicID: 0005

Digest: 垃圾，难看的影响市容，都拿不出手。真心不推荐。三星大边框傻吧。丑是一定的。跟 MIUI 相比差的真的不是一点半点。建议不要，很卡，真心的。

【评价指标】

ROUGE

2.2 用户评论的倾向与方面

在产品营销、网络舆情、民意调查等领域，用户评论具有重要的挖掘价值。本项评测通过影视评论的挖掘，考察对于用户评论的方面提取和倾向性分析能力。该任务包含 3 个子任务，具体如下。

子任务 1: 判断给定影视评论的总体倾向（正面、负面、中立）。

子任务 2: 判断给定影视评论中每个短句所属的方面，包括剧情、导演、演员、人物、配乐、画面、其他等。

子任务 3: 判断给定影视评论中每个短句针对所属方面的倾向性（正面、负面、中立），同时抽取倾向性判断的描述词语，即做出倾向性判断的支撑证据。

其中的“方面”又叫分面、侧面，其含义解释如下：

- **剧情：**包括情节、节奏、一幕场景、剧本等。电影带给观众的主观感受判定为剧情，例如：感人、搞笑等
- **导演：**导演信息

- 演员：包括动画类电影的配音演员
- 人物：人物设计、塑造、性格等，注意演员姓名代指人物的情况
- 配乐：包括主题曲、音效等
- 画面：包括特效、镜头、分镜、摄影、灯光等
- 其他：总体、道具、场景设置、字幕、翻译、背景、政治倾向、细节等其他属性

注意：影视评论中可能存在与影视本身无关的部分，这些短句不属于评测范围。

【提交结果格式】

子任务 1：

Id Run-tag ReviewID Polarity

子任务 2：

Id Run-tag ReviewID SentenceX FacetY

子任务 3：

Id Run-tag ReviewID SentenceX FacetY FacetY_polarity

FacetY_evidence

注意：不同字段以 tab 分割

说明：

Id: 任务编号（此任务编号为2）

Run-tag: 评测参与者编号

ReviewID: 影视评论编号

Polarity: 影视评论的总体倾向性

SentenceX: 影视评论中的短句编号

FacetY: 影视评论中短句所属的方面，包括剧情、导演、演员、人物、配乐、画面、其它等

FacetY_polarity: 影视评论中短句针对所属方面的倾向性

FacetY_evidence: 做出倾向性判断的词语级支撑证据

【任务示例】

测试文本：

<DOC>

<DOCID> 0001 </DOCID>

<TEXT>

<SENTENCE ID="0001_01">"真不忍心说梁朝伟任何负面语言，就这样吧。" </SENTENCE>

<SENTENCE ID="0001_02">"王家卫式的台词也顶不住下午两点的困乏"</SENTENCE>

<SENTENCE ID="0001_03">"在电影院看到睡过去，醒来再接着看"</SENTENCE>

</TEXT>

</DOC>

提交结果:

子任务 1:

2-1 ***** 0001 -1

子任务 2:

2-2 ***** 0001 0001_01 演员

2-2 ***** 0001 0001_02 其它

2-2 ***** 0001 0001_03 其它

子任务 3:

2-3 ***** 0001 0001_01 演员 -1 不忍心 负面

2-3 ***** 0001 0001_02 其它 -1 困乏

2-3 ***** 0001 0001_03 其它 -1

【评价指标】

P、R、F1

2.3 关系分类和关系对抽取

关系抽取是指从文本文档中提取出实体对并识别其具有的关系，目前在知识图谱等系统中具有关键作用。英文的关系抽取研究较早，从 MUC、ACE 到当前的 TAC-KBP 评测会议，已经积累了大量的研究语料。国内虽然不乏关系抽取的顶级学者，但受限于中文关系抽取

语料的匮乏，中文关系抽取的研究进展并不理想。本评测任务希望唤起同仁们对这方面的重视，共同推进中文关系抽取研究的发展。

该评测任务包含两个子任务，具体如下。

子任务 1：给定若干句子，识别出包含实体关系的句子，判断每个句子包含的实体关系类型，不包含任何实体关系的句子无需在结果文件中输出。

子任务 2：给定若干句子，抽取每个句子中包含的特定关系的实体对。不包含任何实体关系的句子无需在结果文件中输出。

本次评测中，限定实体关系类型为如下 10 类（关系名称+关系类型标号）：

人物的出生日期: cr2

人物的出生地: cr4

人物的毕业院校: cr16

人物的配偶: cr20

人物的子女: cr21

组织机构的高管: cr28

组织机构的员工数: cr29

组织机构的创始人: cr34

组织机构的成立时间: cr35

组织机构的总部地点: cr37

【提交结果格式】

子任务 1:

Id Run-tag SentenceID Relationship

子任务 2:

Id Run-tag SentenceID Relationship E1 E2

注意：不同字段以 tab 分割

Id: 任务编号（此任务编号为 3）

Run-tag: 评测参与者编号

SentenceID: 句子 ID

Relationship: 关系类型标号

E1: 实体 1

E2: 实体 2

【任务示例】

测试文本（不同字段以 tab 分割）：

0001 10 月 18 日<e1>桐华</e1>出生于陕西省,早年以陕西省第 9 名的成绩考入<e2>北京大学光华管理学院</e2>

0002 2005 年，创作第一部清穿宫廷小说《步步惊心》。

0003 1901 年与李叔同、黄炎培等同入南洋公学

提交结果：不同字段以 tab 分割

任务 3.1:

3-1 ***** 0001 cr2
3-1 ***** 0001 cr4
3-1 ***** 0001 cr16
3-1 ***** 0003 cr16

任务 3.2：每个关系实体对提交一个结果

3-2 ***** 0001 cr2 桐华 10 月 18 日
3-2 ***** 0001 cr4 桐华 陕西省
3-2 ***** 0001 cr16 桐华 北京大学光华管理学院
3-2 ***** 0003 cr16 李叔同 南洋公学
3-2 ***** 0003 cr16 黄炎培 南洋公学

3. 评测日程

2016.9.5-2016.9.20：标注训练数据与评测数据。

2016.9.12-2016.9.20，参评单位提交报名表，9.20 报名截止评测开始；

2016.9.20：公布评测训练数据；

2016.9.21-2016.10.20：各参评单位开发评测系统；

2016.9.30：公布评测数据与限定资源；

2016.9.30-2016.10.20：各参评单位对参加任务运行测试；

2016.10.20：提交运行结果 Dead Line；

2016.10.21-10.31：专家评判小组对结果进行评判标注；

2016.11.1: 发布评测结果;

2016.10.19-10.31: 各参评单位按照 CCIR2016 的要求与格式撰写评测技术论文;

2016.11.1-2016.11.3: 向 CCIR2016 投稿;

2016.11.4-2016.11.9: 投稿论文按照 CCIR2016 的大会论文标准评审, 享受同其它投稿论文一样的推荐录用机会;

2016.11.9: 公布审稿结果, 印制 notebook 版评测论文集;

2016.11.11-12: 在 CCIR2016 会议上举行评测 workshop, 散发评测论文集, 邀请部分单位宣读论文, 交流经验和技巧, 讨论下一届评测任务。

4. 评测组织

- 评测主办机构

中国中文信息学会信息检索专业委员会

- 评测网址

<http://>

- 评测组织机构

福州大学

北京邮电大学

中国科学院计算技术研究所

- 评测委员会

联合主席：

许洪波 中国科学院计算技术研究所

廖祥文 福州大学

徐蔚然 北京邮电大学

陈 光 北京邮电大学

评测委员：（根据实际参与情况进行遴选）

王素格 山西大学

万小军 北京大学

徐睿峰 哈尔滨工业大学深圳研究生院

王厚峰 北京大学

林鸿飞 大连理工大学

张 奇 复旦大学

刘奕群 清华大学

姚天昉 上海交通大学

黄民烈 清华大学

刘 康 中国科学院自动化研究所