COAE2014评测数据

第六届中文倾向性分析评测（COAE2014）大纲



1.引言

文本倾向性（观点和情感等）分析近几年持续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，前五届中文倾向性分析评测（COAE）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目前亟待解决的问题。因此，在前五届中文倾向性分

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，同时增加

面向新闻的情感关键句抽取任务、跨语言情感分析任务、面向微博的情感词发现任务，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。

与往年评测的一个不同点是，今年的评测技术论文需按照CCIR2014的要求与格式来撰

写，并向CCIR2014投稿，投稿论文参与CCIR2014的大会论文评审，享受同其它投稿论文一

样的录用机会。

2.评测任务设置

第六届COAE2014评测在前五届中文倾向性分析评测的基础上，重点对情感关键句、跨语言情感分析、微博情感新词、微博倾向性、微博观点要素识别进行评测。

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 面向新闻的 情感关键句抽取与判定在给定新闻集合(每篇文章已切成句子)中，判别

每篇文章的情感关键句。

任务2 篇章级 跨语言情感倾向性分析 本任务是对给定多语言篇章级数据集，要求参加系统自动分析多语言篇章级数据集每个情感句的情感倾向性，即（褒义、贬义或者中性）。

任务3 词语级 微博情感新词发现与判定

本任务是对给定大规模的微博句子集，要求参赛系统自动发现新的词语(不在给定的词典以内)，以及每个词语的情感倾向性，即褒义、贬义或者中性。

任务4 句子级 微博观点句识别

在给定的微博句子集合中，判别每个句子的情感倾向性

任务5 要素级 微博观点句评价对象识别

任务5在任务4识别结果的基础上进行，要求找出任务4所识别的每个观点句中观点所针对的评价对象，并对所做评价的倾向性做出判断。

2.1 面向新闻的情感关键句抽取与判定

门户网站、博客和论坛中的新闻性文章很多具有倾向性，倾向性判断对了解社会动态和舆情状况有重要作用。本评测任务侧重于篇章的倾向关键句的抽取、以及对倾向关键句倾向的判定问题，这也是倾向性判断的关键问题之一。

新闻倾向判断中至少有两种倾向。一种是新闻事件本身的正面性，例如自然灾害人员财产损失的报道属于负面新闻，科技、体育和人文方面的进展属于正面新闻。另一种是某新闻报道本身的倾向性，例如“高铁车票打折”事件，有的新闻进行褒奖，有的进行批判。本评测优先选择第二种倾向，即新闻作者本身的倾向性；其次选择新闻事件本身的正面性。

所谓倾向关键句必须能够表达篇章主题的总体倾向性。因此，倾向关键句需要包含两个要素：主题关键词和倾向关键词。主题关键词用来概括篇章的主题；倾向关键词用来概括倾向。

很多新闻报道中会有篇章的倾向句和很多段落倾向句，评测中将优先选择篇章的倾向句。

如果新闻报道中同时存在多个方面的倾向，即该文章综合了多方面的观点，则优先选择最主要的倾向；如果正负两个方面的倾向，并且分量相差不多，则选择两个方面的倾向关键句。此外如果存在多个类似的倾向关键句，则优先选择篇章首尾的句子。

测试文档中的句子已经标定了序号。测试方需要提交每个测试文档的关键句编号和倾向判断结果。

提交结果格式：

Id DocID SenID Run‐tag Polarity

Id: 任务编号（此任务编号为1）

SenID：句子编号

Run‐tag：评测系统编号

Polarity：观点极性，1代表褒义，0代表中性，‐1代表贬义

说明：一行只能包含一个关键句标号，如果一个文章有多个关键句，则提交多行，但一

个文章最多提交两个关键句。没有关键句则没有结果。

注意：不同字段以 tab 分割，下同

2.2 跨语言情感倾向性分析

随着互联网的快速发展和全球化进程的加快，因特网所提供的信息资源呈现出多语言化的特点。全球标准互联网用户调查和分析权威机构Nielsen‐NetRating的调查数据显示，从2000 年到2007年的8年间，全世界各种语言的网络使用增长率达到244.7%。因特网资源的跨语言性及用户对母语和非母语熟悉程度的差异性，不可避免地给用户利用网络信息带来了语言障碍。因此，如何让只会一种语言的用户同时获得多种语言对某个对象的评价倾向性具有重要的应用价值与现实意义。

本任务是对给定多语言篇章级数据集，要求参加系统自动分析多语言篇章级数据集每个情感句的情感倾向性，即褒义、贬义或者中性。因为本任务是一个跨语言任务，所以，本任务只能使用一个语言(中文)的标注语料(规模、领域不做限制)来训练，而对目标语言，只能使用未标注语料(一些目标语言网站的评论评分，也算作近似标签，不建议参赛队使用)。 至于参赛队使用机器翻译，句子对齐，短语对齐，还是可比语料等等，我们均不作限制。

提交结果格式：

Id SenIDRun‐tag Result

Id: 任务编号（此任务编号为2.2）

SenID：句子编号

Run‐tag：评测系统编号

Polarity：观点极性，1代表褒义，0代表中性，‐1代表贬义

例如下面三个句子

<Doc1>Le personnel est d'une très grande gentillesse.</Doc1>

<Doc2> Nous reviendrions volontiers dans cet hôtel très sympathique.</Doc2>

<Doc3> Accueil pas très chaleureux. </Doc3>

输出结果：

2.2 Doc1 xxx 1

2.2 Doc2 xxx 1

2.2 Doc3 xxx ‐1

2.3 微博情感新词发现与判定

微博作为一种社交网站的主要媒体形式，它短小、精悍、快捷，越来越受到人们的喜爱。

人们倾向于从微博上获取资讯、新闻、观点、评论、娱乐等信息，不知不觉间，微博对网络舆情的传播施加了越来越重要的影响。

微博是一种极其活跃且口语化的语言，每天都会有新的富含情感的新词冒出来，因此，发现与识别这些新词对微博的倾向性分析具有重要的意义。

本任务是对给定大规模的微博句子集(千万级规模)，要求参赛系统自动发现新的词语(不在给定的词典以内)，以及每个词语的情感倾向性，即褒义、贬义或者中性。

本任务对词法分析工具不作限制，但建议参赛队伍提前说明一下。

提交结果格式：

Id SenIDRun‐tag Polarity

Id: 任务编号（此任务编号为2.3）

SenID：句子编号

Run‐tag：评测系统编号

New‐word：新词语

Polarity：观点极性，1代表褒义，0代表中性，‐1代表贬义

例如下面四个句子

<Doc1>诺基亚lumia 新款win8系统，用起来很给力哦，很好用</Doc1>

<Doc2>诺基亚的价格策略是....脑残</Doc2>

<Doc3>真坑爹啊，有钱还真的是了不起</Doc3>

输出结果：

2.3 Doc1 xxx 给力 1

2.3 Doc2 xxx 脑残 ‐1

2.3 Doc3 xxx 坑爹 ‐1

2.4 微博倾向性分析

微博是一种口语化的短文本，用词精炼，简洁，存在噪音比如拼写错误，但其中富含观点、倾向与态度。因此，识别微博句子中的倾向性，具有重要的现实意义。

本任务是对给定微博句子集，要求参赛系统自动分析微博句子集中每个句子的情感倾向性，即褒义、贬义或者中性。

提交结果格式：

Id SenIDRun‐tag Polarity

Id: 任务编号（此任务编号为2.4）

SenID：句子编号

Run‐tag：评测系统编号

Polarity：观点极性，1代表褒义，0代表中性，‐1代表贬义

例如下面四个句子

<Doc1>诺基亚lumia 新款win8系统，用起来很流畅哦，很好用</Doc1>

<Doc2>我舍友的诺基亚Lumia竟然摔坏了，就摔了一下</Doc2>

<Doc3>试玩了诺基亚lumia920，感觉浏览器、应用软件的质量比ios、andorid差了好多</Doc3>

<Doc4>不管怎么样，奥迪A6L一直是我的最爱，最爱，最爱</Doc4>

<Doc5>强烈推荐北京银行信用卡。人家北京银行不差钱儿，是偶用过的信用卡中活动最多、礼品最好滴。</Doc5>

输出结果：

2.4 Doc1 xxx 1

2.4 Doc2 xxx ‐1

2.4 Doc3 xxx ‐1

2.4 Doc4 xxx 1

2.4 Doc5 xxx 1

2.5 微博观点句要素抽取

此任务是上一个任务的延续，要求系统从所识别的微博观点句中抽取相应的评价对象，并判别观点句中的倾向性。例如，第一句对于诺基亚lumia 质量进行了分析，我们需要抽取出被评价的产品名、被评价的产品属性以及相对应的观点倾向性。

提交结果格式：

Id SenIDRun‐tag ProductNameFeatureName Polarity

Id: 任务编号（此任务编号为2.5）

SenID：句子编号

Run‐tag：评测系统编号

ProductName: 被评价的产品名

FeatureName：产品属性名

Polarity：观点极性，1代表褒义，0代表中性，‐1代表贬义

例如下面四个句子

<Doc1>诺基亚lumia 新款win8系统，用起来很流畅哦，很好用</Doc1>

<Doc2>我舍友的诺基亚Lumia竟然摔坏了，就摔了一下</Doc2>

<Doc3>试玩了诺基亚lumia920，感觉浏览器、应用软件的质量比ios、andorid差了好多</Doc3>

<Doc4>不管怎么样，奥迪A6L一直是我的最爱，最爱，最爱</Doc4>

<Doc5>强烈推荐北京银行信用卡。人家北京银行不差钱儿，是偶用过的信用卡中活动最多、礼品最好滴。</Doc5>

输出结果应为：

2.5 Doc1 xxx 诺基亚lumia win8系统 1

2.5 Doc2 xxx 诺基亚lumia NULL ‐1

2.5 Doc3 xxx 诺基亚lumia920 浏览器、应用软件 ‐1

2.5 Doc3 xxx ios、andorid 浏览器、应用软件 ‐1

2.5 Doc4 xxx 奥迪A6L NULL 1

2.5 Doc5 xxx 北京银行信用卡 活动 1

2.5 Doc5 xxx 北京银行信用卡 礼品 1

补充说明：

(1) 用顿号、逗号连写的属性与对象可以不分开，如上面的<doc3>等。