

# 中文否定句的情感倾向性分析

张志飞 卫志华 苗夺谦

同济大学  
计算机科学与技术系

第五届中文倾向性分析评测, 2013



# 汇报内容

- 1 引言
- 2 词典构造
- 3 句子的情感倾向识别
- 4 评测结果
- 5 总结



# 汇报内容

- 1 引言
- 2 词典构造
- 3 句子的情感倾向识别
- 4 评测结果
- 5 总结



# 评测任务

第五届中文倾向性分析评测(COAE 2013)共设置了4个任务:

- 基于否定句的句子级观点倾向性分析
- 比较句的识别与要素抽取
- 微博观点句识别
- 微博观点句评价对象识别

否定句的识别非常关键，相对较容易，只参与了该子任务。



# 评测任务

第五届中文倾向性分析评测(COAE 2013)共设置了4个任务:

- 基于否定句的句子级观点倾向性分析
- 比较句的识别与要素抽取
- 微博观点句识别
- 微博观点句评价对象识别

否定句的识别非常关键，相对较容易，只参与了该子任务。



# 解决思路

## 词典、规则和统计相结合

### ● 词典

- 情感词表：上下文无关词语的情感
- 否定词表：情感极性转向
- 程度词表：情感强度增强或减弱
- 连词词表：融合子句关系的情感求和
- 规则：强语义模糊性词语“好”
- 统计：上下文相关词语<sup>1</sup>的情感

<sup>1</sup>限定常用的单字节形容词，如“高”、“低”、“大”、“小”等



# 解决思路

## 词典、规则和统计相结合

### ● 词典

- 情感词表：上下文无关词语的情感
- 否定词表：情感极性转向
- 程度词表：情感强度增强或减弱
- 连词词表：融合子句关系的情感求和
- 规则：强语义模糊性词语“好”
- 统计：上下文相关词语<sup>1</sup>的情感



<sup>1</sup>限定常用的单字节形容词，如“高”、“低”、“大”、“小”等

# 解决思路

## 词典、规则和统计相结合

### ● 词典

- 情感词表：上下文无关词语的情感
- 否定词表：情感极性转向
- 程度词表：情感强度增强或减弱
- 连词词表：融合子句关系的情感求和
- 规则：强语义模糊性词语“好”
- 统计：上下文相关词语<sup>1</sup>的情感

<sup>1</sup>限定常用的单字节形容词，如“高”、“低”、“大”、“小”等





# 解决思路

## 词典、规则和统计相结合

### ● 词典

- 情感词表：上下文无关词语的情感
- 否定词表：情感极性转向
- 程度词表：情感强度增强或减弱
- 连词词表：融合子句关系的情感求和

### ● 规则：强语义模糊性词语“好”

### ● 统计：上下文相关词语<sup>1</sup>的情感



<sup>1</sup>限定常用的单字节形容词，如“高”、“低”、“大”、“小”等

# 解决思路

## 词典、规则和统计相结合

### ● 词典

- 情感词表：上下文无关词语的情感
- 否定词表：情感极性转向
- 程度词表：情感强度增强或减弱
- 连词词表：融合子句关系的情感求和

### ● 规则：强语义模糊性词语“好”

### ● 统计：上下文相关词语<sup>1</sup>的情感



<sup>1</sup>限定常用的单字节形容词，如“高”、“低”、“大”、“小”等

# 解决思路

## 词典、规则和统计相结合

### ● 词典

- 情感词表：上下文无关词语的情感
- 否定词表：情感极性转向
- 程度词表：情感强度增强或减弱
- 连词词表：融合子句关系的情感求和

### ● 规则：强语义模糊性词语“好”

### ● 统计：上下文相关词语<sup>1</sup>的情感

<sup>1</sup>限定常用的单字节形容词，如“高”、“低”、“大”、“小”等



# 解决思路

## 词典、规则和统计相结合

### ● 词典

- 情感词表：上下文无关词语的情感
- 否定词表：情感极性转向
- 程度词表：情感强度增强或减弱
- 连词词表：融合子句关系的情感求和

### ● 规则：强语义模糊性词语“好”

### ● 统计：上下文相关词语<sup>1</sup>的情感

<sup>1</sup>限定常用的单字节形容词，如“高”、“低”、“大”、“小”等。



# 汇报内容

- 1 引言
- 2 词典构造
- 3 句子的情感倾向识别
- 4 评测结果
- 5 总结



# 情感词表

情感词语是情感倾向性分析的基础，用符号 $word$ 表示。

规模：**28567**个

来源：情感词汇本体<sup>2</sup>、学生褒贬义词典<sup>3</sup>、日常用语

词语的情感值计算(强度五级制转化为[1,2]):

$$Sentiment(word) = Polarity(word) \times 0.2 \times Intensity(word) \quad (1)$$

$Polarity$ 代表极性,  $Polarity(word) \in \{-1, 0, 1\}$

$Intensity$ 代表强度,  $Intensity(word) \in \{1, 2, 3, 4, 5\}$

<sup>2</sup>徐琳宏, 林鸿飞, 潘宇, 等. 情感词汇本体的构造. 情报学报, 2008, 27(2): 180–185.

<sup>3</sup><http://tjzhifei.github.io/resources/bbycd.zip>



# 否定词表

否定词的出现会改变情感极性，用符号 $neg$ 表示。

规模：65个

来源：文献<sup>4</sup>、手工

Table 1： 否定词表

不	没	非	别	无
不是	不能	不会	并未	难以
没有	从不	决不	无法	毫不
从未	未曾	才怪	不像	尚未
不怎么	不见得	说不上	算不上	木有

<sup>4</sup>唐都钰, 石秋慧, 等. HITIRSYS: COAE2012 情感分析系统. 第四届中文倾向性分析评测论文集, 2012.



# 程度词表

程度词会加强或者减弱情感，但不改变情感倾向，用符号 $deg$ 表示。

规模：**140**个

来源：HowNet<sup>5</sup>中的“程度级别词语(中文)”

权重：**5**(“极其”意义) **4**(“很”意义) **3**(“较”意义) **2**(“稍微”意义)

Table 2 : 程度词表

词语	权重	词语	权重
非常	5	更加	3
极其	5	越来越	3
特别	4	略微	2
不是一般	4	挺	2

程度词的程度值的计算：

$$Degree(deg) = 0.25 \times (3 + Weight(deg)) \quad (2)$$

$Weight$ 表示程度权重， $Weight(deg) \in \{2, 3, 4, 5\}$

<sup>5</sup><http://www.keenage.com/>





# 连词词表

不同类型的从句表达情感的重要程度不同(转折> 让步>假设= 目的=条件)。

规模: **50**个, 其中转折词12个, 让步词7个

来源: 手工

Table 3 : 连词词表

词语	类型	词语	类型
尽管	让步	就算	让步
但是	转折	不过	转折
如果	假设	要是	假设
为了	目的	以便	目的
只有	条件	除非	条件



# 汇报内容

- 1 引言
- 2 词典构造
- 3 句子的情感倾向识别
- 4 评测结果
- 5 总结



# 预处理

- 分词和词性标注：ICTCLAS<sup>6</sup>
- 用户词典：影响情感判断的未登录词，如“幸福使者”、“卓越”
- 切分子句：根据标点符号，。？！；

<sup>6</sup><http://ictclas.nlpir.org/>



# 预处理

- 分词和词性标注：ICTCLAS<sup>6</sup>
- 用户词典：影响情感判断的未登录词，如“幸福使者”、“卓越”
- 切分子句：根据标点符号，。？！；

<sup>6</sup><http://ictclas.nlpir.org/>



# 预处理

- 分词和词性标注：ICTCLAS<sup>6</sup>
- 用户词典：影响情感判断的未登录词，如“幸福使者”、“卓越”
- 切分子句：根据标点符号，。？！；

<sup>6</sup><http://ictclas.nlpir.org/>



# 汇报内容

## 1 引言

## 2 词典构造

## 3 句子的情感倾向识别

- PartA: 子句的情感值计算
- PartB: “好”字的情感值计算
- PartC: 上下文相关词语的情感值计算
- PartD: 句子的情感值计算

## 4 评测结果

## 5 总结



# 总体思路

假设：

- 子句表达的情感较为单一 (子句长度较短)
- 含有多个情感词时以最后一个为准 (符合中国人的表达思维方式)

只考虑与情感词距离为3之内的否定词和程度词

搭配	子句的情感值	举例
$deg + word$	$Sentiment(subSen) = Degree(deg) \times Sentiment(word)$	很差
$neg + word$	$Sentiment(subSen) = -Sentiment(word)$	不差
$neg + neg + word$	$Sentiment(subSen) = Sentiment(word)$	不是不好
$deg + neg + word$	$Sentiment(subSen) = -Degree(deg) \times Sentiment(word)$	非常不厚道
$neg + deg + word$	$Sentiment(subSen) = -Sentiment(word) / Degree(deg)$	不是很差

Table 4 : 情感词、否定词和程度词搭配<sup>7</sup>

<sup>7</sup>该表只列出了否定词和程度词前置情况



# 总体思路

假设：

- 子句表达的情感较为单一(子句长度较短)
- 含有多个情感词时以最后一个为准(符合中国人的表达思维方式)

只考虑与情感词距离为3之内的否定词和程度词

搭配	子句的情感值	举例
$deg + word$	$Sentiment(subSen) = Degree(deg) \times Sentiment(word)$	很差
$neg + word$	$Sentiment(subSen) = -Sentiment(word)$	不差
$neg + neg + word$	$Sentiment(subSen) = Sentiment(word)$	不是不好
$deg + neg + word$	$Sentiment(subSen) = -Degree(deg) \times Sentiment(word)$	非常不厚道
$neg + deg + word$	$Sentiment(subSen) = -Sentiment(word) / Degree(deg)$	不是很差

Table 4 : 情感词、否定词和程度词搭配<sup>7</sup>

<sup>7</sup>该表只列出了否定词和程度词前置情况





# 关键问题

计算 $Sentiment(word)$ :

- 若 $word$ 在情感词典中且具有唯一极性，则根据Eq.1直接计算；
- 若 $word$ 是强语义模糊性词语“好”，则根据PartB的方法计算；
- 若 $word$ 是上下文相关的情感词语，则根据PartC的方法计算。



# 关键问题

计算 $Sentiment(word)$ :

- 若 $word$ 在情感词典中且具有唯一极性，则根据Eq.1直接计算；
- 若 $word$ 是强语义模糊性词语“好”，则根据PartB的方法计算；
- 若 $word$ 是上下文相关的情感词语，则根据PartC的方法计算。



# 关键问题

计算 $Sentiment(word)$ :

- 若 $word$ 在情感词典中且具有唯一极性，则根据Eq.1直接计算；
- 若 $word$ 是强语义模糊性词语“好”，则根据PartB的方法计算；
- 若 $word$ 是上下文相关的情感词语，则根据PartC的方法计算。



# 汇报内容

## 1 引言

## 2 词典构造

## 3 句子的情感倾向识别

- PartA: 子句的情感值计算
- PartB: “好”字的情感值计算
- PartC: 上下文相关词语的情感值计算
- PartD: 句子的情感值计算

## 4 评测结果

## 5 总结



# “好”字的问题

## 例1 还好退了

ICTCLAS: 还/d 好/a 退/v 了/y ☹

HIT LTP: 还/d 好/a 退/v 了/u ☹

FudanNLP: 还/副词 好/副词 退/动词 了/时态词 ☹

Stanford Parser: advmod(退,还好) ☹

## 例2 设置好不用20分钟

ICTCLAS: 设置/v 好/a 都/d 不/d 用/v 20/m 分钟/q ☹

HIT LTP: 设置/v 好/a 都/d 不/d 用/v 20/m 分钟/q ☹

FudanNLP: 设置/动词 好/趋向动词 都/副词 不/副词 用/动词 20/ 数词 分钟/量词 ☹

Stanford Parser: dep(设置,好) ☹

## 例3 维修站的服务好差

ICTCLAS: 维修/v 站/v 的/u 服务/v 好/a 差/a ☹

HIT LTP: 维修站/n 的/u 服务/v 好/a 差/a ☹

FudanNLP: 维修站/名词 的/结构助词 服务/名词 好差/名词 ☹

Stanford Parser: nsubj(好,服务) root(ROOT,好) comod(好,差) ☹

本文仅仅使用ICTCLAS，然后通过规则确定“好”是否表达情感！

# “好”字的问题

## 例1 还好退了

ICTCLAS: 还/d 好/a 退/v 了/y ☹

HIT LTP: 还/d 好/a 退/v 了/u ☹

FudanNLP: 还/副词 好/副词 退/动词 了/时态词 ☹

Stanford Parser: advmod(退,还好) ☹

## 例2 设置好不用20分钟

ICTCLAS: 设置/v 好/a 都/d 不/d 用/v 20/m 分钟/q ☹

HIT LTP: 设置/v 好/a 都/d 不/d 用/v 20/m 分钟/q ☹

FudanNLP: 设置/动词 好/趋向动词 都/副词 不/副词 用/动词 20/ 数词 分钟/量词 ☹

Stanford Parser: dep(设置,好) ☹

## 例3 维修站的服务好差

ICTCLAS: 维修/v 站/v 的/u 服务/v 好/a 差/a ☹

HIT LTP: 维修站/n 的/u 服务/v 好/a 差/a ☹

FudanNLP: 维修站/名词 的/结构助词 服务/名词 好差/名词 ☹

Stanford Parser: nsubj(好,服务) root(ROOT,好) comod(好,差) ☹

本文仅仅使用ICTCLAS，然后通过**规则**确定“好”是否表达情感！

# “好”的释义

“好”具有很强的语义模糊性<sup>8</sup>。

编号	释义	极性	编号	释义	极性
1	表示一切美好的事物	褒义	9	应该；可以	中性
2	优点多的；使人满意的	褒义	10	便于；以便	中性
3	合宜；妥当	褒义	11	表示程度深	中性
4	友爱；和睦	褒义	12	表示多或久	中性
5	健康；痊愈	褒义	13	表示完成或达到完善的地步	中性
6	表示使人满意的性质	褒义	14	表示结束语气或者话语衔接	中性
7	表示容易	褒义	15	用于套语	中性
8	表示赞许、同意语气	褒义	16	反话，表示不满意	贬义

<sup>8</sup>孙秋秋. “好”在语义上的模糊性与确定性. 辽宁大学学报(哲学社会科学版), 1982, (1): 70-76.



# “好”的分布

语料库：第四届中文倾向性分析评测<sup>9</sup>COAE2012 任务1

Table 5： COAE2012任务1语料库统计

领域	句子总数	含有“好”字的句子总数
电子	1200	309
汽车	1200	525



<sup>9</sup>刘康, 王素格, 等. 第四届中文倾向性分析评测总体报告. 第四届中文倾向性分析评测论文集, 2012, 1-32.





# 9条规则

## 规则A (右端规则)

若“好”的右边是标点符号或者句子结束符号，“好”为褒义。

## 规则B (程度副词相邻规则)

若“好”与程度副词(除“最”)相邻，“好”为褒义。

## 规则C (“得的啊”相邻规则)

若“好”与“得”、“的”或“啊”相邻，“好”为褒义。

## 规则D (形容词右邻规则)

若“好”的右边是形容词，“好”等于“很”。

## 规则E (名词相邻规则)

若“好”与名词相邻，“好”为褒义。

## 9条规则(续)

### 规则F (代词左邻规则)

若“好”的左边是代词，“好”为褒义。

### 规则G (“不”右邻规则)

若“好”的右边是“不”，当再右边的词在“好不AP”表<sup>a</sup>中，根据该表确定“好不”等于“很”还是“不”，否则不考虑“好”。

<sup>a</sup><http://tjzhifei.github.io/resources/haobuap.rtf>

### 规则H (动词右邻规则)

若“好”的右边是动词，当再右边是标点符号或者子句结束符号，“好”为褒义，否则不考虑“好”。

### 规则I (动词左邻规则)

若“好”的左边是动词，不考虑“好”。

# 举例说明

## 例1 还好退了

ICTCLAS: 还/d 好/a 退/v 了/y ☺

匹配规则H:  $Sentiment(\text{好}) = 0$  ☺

## 例2 设置好不用20分钟

ICTCLAS: 设置/v 好/a 都/d 不/d 用/v 20/m 分钟/q ☺

匹配规则I:  $Sentiment(\text{好}) = 0$  ☺

## 例3 维修站的服务好差

ICTCLAS: 维修/v 站/v 的/u 服务/v 好/a 差/a ☺

匹配规则D:  $Sentiment(\text{好差}) = Degree(\text{好}) \times Sentiment(\text{差})$  ☺

通过匹配规则能够很好地确定“好”是否表达情感!



# 举例说明

## 例1 还好退了

ICTCLAS: 还/d 好/a 退/v 了/y ☺

匹配规则H:  $Sentiment(\text{好}) = 0$  ☺

## 例2 设置好不用20分钟

ICTCLAS: 设置/v 好/a 都/d 不/d 用/v 20/m 分钟/q ☺

匹配规则I:  $Sentiment(\text{好}) = 0$  ☺

## 例3 维修站的服务好差

ICTCLAS: 维修/v 站/v 的/u 服务/v 好/a 差/a ☺

匹配规则D:  $Sentiment(\text{好差}) = Degree(\text{好}) \times Sentiment(\text{差})$  ☺

通过匹配规则能够很好地确定“好”是否表达情感!



# 汇报内容

## 1 引言

## 2 词典构造

## 3 句子的情感倾向识别

- PartA: 子句的情感值计算
- PartB: “好”字的情感值计算
- PartC: 上下文相关词语的情感值计算
- PartD: 句子的情感值计算

## 4 评测结果

## 5 总结



# 上下文相关词语

## 例子

褒义：这件商品的性价比很高

贬义：这件商品的性价比很低

褒义：这款车的油耗很低

贬义：这款车的油耗很高

“高”和“低”在不同的上下文中表达不同的情感倾向。

“高”和“低”是一组反义词对。



# 基本思路

假设：在同一上下文中，互为反义的两个词语表达相反的情感倾向。

利用反义词对识别上下文相关词语<sup>10</sup>的情感倾向：

- 上下文信息抽取
- 计算情感后验概率
- 情感倾向识别

<sup>10</sup> 目前只处理了“大、小、高、低、快、慢、长、短”



## 上下文信息抽取

上下文主要指评价对象的特征，记作 $f$ 。

选取与情感词距离为3以内的名词且出现次数较多的作为特征。

以COAE2012任务1语料库为例：





# 计算情感后验概率

基于贝叶斯定理，给定特征后词语的情感倾向后验概率为：

$$P(w = 1|f) = \frac{P(f|w = 1)P(w = 1)}{\sum_{u=-1}^1 P(f|w = u)P(w = u)} \quad (3)$$

$$P(w = -1|f) = \frac{P(f|w = -1)P(w = -1)}{\sum_{u=-1}^1 P(f|w = u)P(w = u)} \quad (4)$$

其中：

$$P(w = u) = \frac{\text{count}(w = u) + 1}{\sum_{v=-1}^1 [\text{count}(w = v) + 1]} \quad (5)$$

$$P(f|w = u) = \begin{cases} \frac{\text{count}(f, w=u)+1}{\text{count}(w=u)+1} & \text{count}(w = u) \neq 0 \\ 0.001 & \text{count}(w = u) = 0 \end{cases} \quad (6)$$



# 情感倾向识别

## 两条准则：

- **准则1：** 给定特征下，如果**反义词对的两个词语均有后验概率**，当某个词语为褒义的后验概率大于(或小于)为贬义的后验概率且它的反义词为褒义的后验概率小于(或大于)为贬义的后验概率，则在出现该特征时，这个词语为褒义(或贬义)，同时它的反义词为贬义(或褒义)；

$$Polarity(w_1|f) = \begin{cases} 1 & P(w_1 = 1|f) > P(w_1 = -1|f) \wedge P(w_2 = 1|f) < P(w_2 = -1|f) \\ -1 & P(w_1 = 1|f) < P(w_1 = -1|f) \wedge P(w_2 = 1|f) > P(w_2 = -1|f) \end{cases} \quad (7)$$

- **准则2：** 给定特征下，如果**反义词对只有一个情感词语有后验概率**，当这个词语为褒义的后验概率减去为贬义的后验概率大于等于(或小于)0.5，则在出现该特征时，这个词语为褒义(或贬义)，同时它的反义词为贬义(或褒义)。

$$Polarity(w_1|f) = \begin{cases} 1 & P(w_1 = 1|f) - P(w_1 = -1|f) \geq 0.5 \\ -1 & P(w_1 = 1|f) - P(w_1 = -1|f) < 0.5 \end{cases} \quad (8)$$

互为反义的词语的情感倾向为： $Polarity(w_2|f) = -Polarity(w_1|f)$



## 情感倾向识别

两条准则：

- **准则1:** 给定特征下, 如果反义词对的两个词语均有后验概率, 当某个词语为褒义的后验概率大于(或小于)为贬义的后验概率且它的反义词为褒义的后验概率小于(或大于)为贬义的后验概率, 则在出现该特征时, 这个词语为褒义(或贬义), 同时它的反义词为贬义(或褒义);

$$Polarity(w_1|f) = \begin{cases} 1 & P(w_1 = 1|f) > P(w_1 = -1|f) \wedge P(w_2 = 1|f) < P(w_2 = -1|f) \\ -1 & P(w_1 = 1|f) < P(w_1 = -1|f) \wedge P(w_2 = 1|f) > P(w_2 = -1|f) \end{cases} \quad (7)$$

- **准则2:** 给定特征下, 如果反义词对只有一个情感词语有后验概率, 当这个词语为褒义的后验概率减去为贬义的后验概率大于等于(或小于)0.5, 则在出现该特征时, 这个词语为褒义(或贬义), 同时它的反义词为贬义(或褒义)。

$$Polarity(w_1|f) = \begin{cases} 1 & P(w_1 = 1|f) - P(w_1 = -1|f) \geq 0.5 \\ -1 & P(w_1 = 1|f) - P(w_1 = -1|f) < 0.5 \end{cases} \quad (8)$$

互为反义的词语的情感倾向为:  $Polarity(w_2|f) = -Polarity(w_1|f)$



# 举例说明

$f$	$w$	$L$	$f$	$w$	$L$	$f$	$w$	$L$
价格	高	-1	噪音	大	-1	速度	慢	-1
价格	高	-1	噪音	大	-1	速度	慢	0
价格	高	-1	噪音	小	1	空间	大	1
价格	高	0	噪音	小	1	空间	大	1
价格	低	1	油耗	低	1	空间	大	1
价格	低	0	油耗	低	1	空间	大	0
质量	高	1	油耗	低	1	空间	大	-1
质量	高	1	速度	快	1	空间	小	-1
质量	高	0	速度	快	1	空间	小	-1
噪音	大	1	速度	快	1	空间	小	0

$P(\text{低}=1|\text{价格})=0.4999$   $P(\text{低}=-1|\text{价格})=0.0002$   $P(\text{高}=1|\text{价格})=0.1429$   $P(\text{高}=-1|\text{价格})=0.5714$   
 $P(\text{小}=1|\text{噪音})=0.6000$   $P(\text{小}=-1|\text{噪音})=0.2000$   $P(\text{大}=1|\text{噪音})=0.3333$   $P(\text{大}=-1|\text{噪音})=0.5000$   
 $P(\text{慢}=1|\text{速度})=0.0002$   $P(\text{慢}=-1|\text{速度})=0.4999$   $P(\text{快}=1|\text{速度})=0.9995$   $P(\text{快}=-1|\text{速度})=0.0002$   
 $P(\text{小}=1|\text{空间})=0.1667$   $P(\text{小}=-1|\text{空间})=0.5000$   $P(\text{大}=1|\text{空间})=0.5000$   $P(\text{大}=-1|\text{空间})=0.2500$   
 $P(\text{高}=1|\text{质量})=0.5000$   $P(\text{高}=-1|\text{质量})=0.1667$   $P(\text{低}=1|\text{油耗})=0.7998$   $P(\text{低}=-1|\text{油耗})=0.0002$

$\text{Polarity}(\text{高}|\text{价格})=1$   $\text{Polarity}(\text{低}|\text{价格})=-1$   $\text{Polarity}(\text{快}|\text{速度})=1$   $\text{Polarity}(\text{慢}|\text{速度})=-1$   
 $\text{Polarity}(\text{低}|\text{油耗})=1$   $\text{Polarity}(\text{高}|\text{油耗})=-1$   $\text{Polarity}(\text{大}|\text{空间})=1$   $\text{Polarity}(\text{小}|\text{空间})=-1$   
 $\text{Polarity}(\text{大}|\text{噪音})=-1$   $\text{Polarity}(\text{小}|\text{噪音})=-1$



# 汇报内容

## 1 引言

## 2 词典构造

## 3 句子的情感倾向识别

- PartA: 子句的情感值计算
- PartB: “好”字的情感值计算
- PartC: 上下文相关词语的情感值计算
- PartD: 句子的情感值计算

## 4 评测结果

## 5 总结



# 子句标记

结合连词类型、标点符号和否定词等，给定子句如下标记：

标记	解释	标记	解释
Single-Polar	子句只含有褒义或者贬义情感	Con-Polar	子句含有褒义或者贬义情感，且包含让步连词
Single-Con	子句只含有让步连词	Tra-Polar	子句含有褒义或者贬义情感，且包含转折连词
Single-Tra	子句只含有转折连词	Other-Polar <sup>11</sup>	子句含有褒义或者贬义情感，且包含其余类型的连词
Single-Neg	子句只含有否定词	Hope-Polar <sup>12</sup>	子句同时表达了一种主观愿望
Single-Punc	子句只含有标点符号	None	子句不出现上述任何情况

<sup>11</sup> 不考虑目的、假设或者条件连词所在句子的情感倾向

<sup>12</sup> 不考虑主观愿望词语所在句子的情感倾向



# 修正子句的情感值

当出现让步和转折情况时，根据下面4种情况修正子句的情感值：

- 当前子句为Single-Tra，如果前一个子句为Single-Polar，将前一个子句的情感值置0，同时如果存在后一个子句为None或者不存在后续子句，那么当前子句的情感值为前一个子句情感值的相反数；<sup>13</sup>
- 当前子句为Single-Con，如果后一个子句为Single-Polar，那么将后一个子句的情感值置0；
- 当前子句为Tra-Polar，如果前一个子句为Single-Polar，那么将前一个子句的情感值置0；
- 当前子句为Con-Polar，则将当前子句的情感值置0。

<sup>13</sup>E.g., 他人不错，但是我很他。



# 计算句子情感极性

将所有子句的情感值累加求和得到句子的情感值：

$$Sentiment(sentence) = \sum_{i=1}^m Sentiment(subSen_i) \quad (9)$$

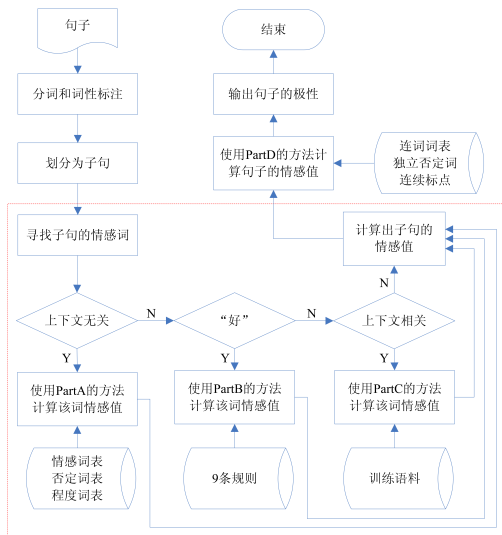
句子的情感极性为：

$$Polarity(sentence) = \begin{cases} 1 & Sentiment(sentence) > 0 \\ -1 & Sentiment(sentence) < 0 \\ -1 & Sentiment(sentence) = 0 \wedge \\ & (count(Single-Neg) > 2 \vee count(Single-Punc) > 1) \\ 0 & \text{otherwise} \end{cases} \quad (10)$$





# 系统流程图



# 汇报内容

- 1 引言
- 2 词典构造
- 3 句子的情感倾向识别
- 4 评测结果**
- 5 总结



# 电子和汽车领域的评测结果

		电子领域			汽车领域		
		TJOMS	Average	Best	TJOMS	Average	Best
正面	<i>Pre</i>	0.6219	0.6243	0.7407	0.3333	0.3302	0.7473
	<i>Rec</i>	0.7811	0.7208	0.8497	<b>0.6575</b>	0.4392	0.6575
	<i>F<sub>1</sub></i>	0.6924	0.6580	0.7595	0.4424	0.3780	0.5666
中性	<i>Pre</i>	0.3622	0.3567	0.5000	<b>0.5732</b>	0.4519	0.5732
	<i>Rec</i>	0.2233	0.2233	0.5252	0.4327	0.3630	0.9279
	<i>F<sub>1</sub></i>	0.2763	0.2750	0.4671	0.4932	0.3969	0.6220
负面	<i>Pre</i>	0.5780	0.5061	0.7669	0.6805	0.6193	0.9060
	<i>Rec</i>	0.5510	0.5845	0.9096	0.5970	0.5945	0.9254
	<i>F<sub>1</sub></i>	<u>0.5642</u>	0.5490	0.6425	<u>0.6360</u>	0.6083	0.6818
<i>Accu</i>		0.5675	0.5388	0.64	0.5492	0.5167	0.5942

汽车领域的正面*Rec*和中性*Pre*取得最好结果，但是电子领域不明显。  
汽车领域和电子领域的负面*F<sub>1</sub>*值均排名靠前。

# 宏平均和微平均评测结果

		宏平均			微平均		
		TJOMS	Average	Best	TJOMS	Average	Best
正面	<i>Pre</i>	0.4776	0.4809	0.7409	0.5222	0.5343	0.7408
	<i>Rec</i>	<b>0.7193</b>	0.6158	0.7193	<u>0.7500</u>	0.6806	0.7792
	$F_1$	0.5740	0.5255	0.6744	0.6157	0.5694	0.7191
中性	<i>Pre</i>	0.4677	0.4082	0.5002	<u>0.4922</u>	0.4156	0.5102
	<i>Rec</i>	0.3280	0.2994	0.7253	0.3420	0.3079	0.7520
	$F_1$	0.3856	0.3488	0.5512	0.4035	0.3603	0.5653
负面	<i>Pre</i>	0.6293	0.5675	0.8365	0.6414	0.5784	0.8406
	<i>Rec</i>	0.5740	0.5777	0.9175	0.5803	0.5809	0.9197
	$F_1$	<u>0.6004</u>	0.5826	0.6467	<u>0.6093</u>	0.5822	0.6494
<i>Accu</i>		0.5583	0.5185	0.6154	0.5583	0.5185	0.6154

正面*Rec*、中性*Pre*和负面 $F_1$ 值均取得了较好的结果。  
正面*Pre*和负面*Rec*稍微低于平均值。

# 系统排名

该子任务共有国内13个单位参与，排名前六的系统如下：

排序	系统名称	<i>Accu</i>
1	DUTIR	0.6154
2	CUCSas	0.5817
3	SCIR	0.5646
4	TJOMS	0.5583
5	PRIS	0.5471
6	WHU	0.5296

本文的系统排名第四。



# 系统排名

该子任务共有国内13个单位参与，排名前六的系统如下：

排序	系统名称	<i>Accu</i>
1	DUTIR	0.6154
2	CUCSas	0.5817
3	SCIR	0.5646
4	TJOMS	0.5583
5	PRIS	0.5471
6	WHU	0.5296

本文的系统排名第四。



# 汇报内容

- 1 引言
- 2 词典构造
- 3 句子的情感倾向识别
- 4 评测结果
- 5 总结



# 总结

参加了COAE2013的子任务1，即中文否定句的情感倾向性分析。

- 词典、规则 and 统计相结合的方法
- 词典主要用于无歧义情感词的情感分析
- 规则用于强语义模糊性词语“好”的情感分析
- 统计用于上下文相关情感词的情感分析
- 融合子句关系的句子情感分析
- 考虑独立否定词和连续标点的作用

本文的方法不局限于否定句，可以辅助任何情感分析任务。





# 总结

参加了COAE2013的子任务1，即中文否定句的情感倾向性分析。

- 词典、规则和统计相结合的方法
- 词典主要用于无歧义情感词的情感分析
- 规则用于强语义模糊性词语“好”的情感分析
- 统计用于上下文相关情感词的情感分析
- 融合子句关系的句子情感分析
- 考虑独立否定词和连续标点的作用

本文的方法不局限于否定句，可以辅助任何情感分析任务。



# 总结

参加了COAE2013的子任务1，即中文否定句的情感倾向性分析。

- 词典、规则和统计相结合的方法
- 词典主要用于无歧义情感词的情感分析
- 规则用于强语义模糊性词语“好”的情感分析
- 统计用于上下文相关情感词的情感分析
- 融合子句关系的句子情感分析
- 考虑独立否定词和连续标点的作用

本文的方法不局限于否定句，可以辅助任何情感分析任务。



# 总结

参加了COAE2013的子任务1，即中文否定句的情感倾向性分析。

- 词典、规则和统计相结合的方法
- 词典主要用于无歧义情感词的情感分析
- 规则用于强语义模糊性词语“好”的情感分析
- 统计用于上下文相关情感词的情感分析
- 融合子句关系的句子情感分析
- 考虑独立否定词和连续标点的作用

本文的方法不局限于否定句，可以辅助任何情感分析任务。



# 总结

参加了COAE2013的子任务1，即中文否定句的情感倾向性分析。

- 词典、规则和统计相结合的方法
- 词典主要用于无歧义情感词的情感分析
- 规则用于强语义模糊性词语“好”的情感分析
- 统计用于上下文相关情感词的情感分析
- 融合子句关系的句子情感分析
- 考虑独立否定词和连续标点的作用

本文的方法不局限于否定句，可以辅助任何情感分析任务。



# 总结

参加了COAE2013的子任务1，即中文否定句的情感倾向性分析。

- 词典、规则和统计相结合的方法
- 词典主要用于无歧义情感词的情感分析
- 规则用于强语义模糊性词语“好”的情感分析
- 统计用于上下文相关情感词的情感分析
- 融合子句关系的句子情感分析
- 考虑独立否定词和连续标点的作用

本文的方法不局限于否定句，可以辅助任何情感分析任务。



# 总结

参加了COAE2013的子任务1，即中文否定句的情感倾向性分析。

- 词典、规则和统计相结合的方法
- 词典主要用于无歧义情感词的情感分析
- 规则用于强语义模糊性词语“好”的情感分析
- 统计用于上下文相关情感词的情感分析
- 融合子句关系的句子情感分析
- 考虑独立否定词和连续标点的作用

本文的方法不局限于否定句，可以辅助任何情感分析任务。



# 总结

参加了COAE2013的子任务1，即中文否定句的情感倾向性分析。

- 词典、规则和统计相结合的方法
- 词典主要用于无歧义情感词的情感分析
- 规则用于强语义模糊性词语“好”的情感分析
- 统计用于上下文相关情感词的情感分析
- 融合子句关系的句子情感分析
- 考虑独立否定词和连续标点的作用

本文的方法不局限于否定句，可以辅助任何情感分析任务。



# 致谢

- COAE2013组委会(本文使用的测试数据及评测结果)
- COAE2012组委会(本文使用COAE2012作为训练语料)
- ICTCLAS(本文分词和词性标注的工具)
- 情感词汇本体(本文情感词典的主要来源)
- HIT LTP、FudanNLP、HowNet等





# Thank you !

## Q & A?

Email: [tjzhifei@163.com](mailto:tjzhifei@163.com)

Weibo: @同济志飞

