

# 一种基于情绪表达与情绪认知分离的新型情绪词典

徐睿峰, 邹承天, 郑燕珍, 徐军, 桂林, 刘滨, 王晓龙

哈尔滨工业大学深圳研究生院, 广东省 深圳市 518055

E-mail: xuruifeng@hitsz.edu.cn

**摘要:** 目前的情绪词典通常对情绪词语进行情绪类别和强度的标注, 但缺乏对词语的情绪表达和情绪认知进行区分的能力。同时, 直接在词语条目上进行标注经常由于词语的语义歧义导致情绪标注结果的歧义。本文在对个体情绪产生和迁移机制进行分析的基础上, 建立了基于“刺激认知-反射表达”文本情绪计算框架。并在此框架下对情绪词语的功能和特性进行分析, 探索了一种新型情绪词典建设方法。首先, 引入HowNet提供的词语语义信息, 将同一词语转变为不同语义的多个词条进行标注减少情绪标注歧义。其次, 将词语的情绪表达结果和情绪认知结果加以区分, 分别标注从不同角度观测到的词条情绪类别和强度。同时对词语的情绪表达和情绪认知类型进行了细化分类。最终初步构建出一个具有清晰框架、丰富情绪信息和较低歧义的新型情绪词典。

**关键词:** 情绪词典; 情绪认知; 情绪表达; 词语语义

## A New Emotion Dictionary based on the Distinguish of Emotion

### Expression and Emotion Cognition

Ruifeng Xu, Chengtian Zou, Yanzhen Zheng, Jun Xu, Lin Gui, Bing Liu, Xiaolong Wang

Harbin Institute of Technology Shenzhen Graduate School, Shenzhen, Guangdong 518055, China

E-mail: xuruifeng@hitsz.edu.cn

**Abstract:** The current emotion dictionary usually annotates the categories and strength of emotion words, but lack of the capacity to distinguish emotional expression and emotional cognitive. Meanwhile, the direct annotation on the words leads to emotion annotation ambiguities caused by the word sense ambiguities. Based on the analysis on the generation and migration mechanisms of individual emotions, this papers proposes a text emotion computing framework based on "cognitive stimulation - reflective expression" mechanism. Under this framework, we explores the construction strategy of a new emotion dictionary through analyze the function and characteristics of emotion words. Firstly, we introduce the part-of-speech and word sense information provided by HowNet for separating one word to multiple entries corresponding to different senses in order to reduce emotional annotation ambiguity. Secondly, the emotion expression and emotion cognition results by the same entry are distinguished. The emotion categories and corresponding strength are annotated from different aspects, respectively. Meanwhile, the types of emotion expression and emotion cognition are refined annotated, respectively. Finally, a preliminary new emotion dictionary is constructed with a clear framework, rich emotional knowledge and low ambiguity.

**Key words:** emotion dictionary; emotion cognition; emotion expression; word sense

## 1 引言

情绪是人类的多种感觉、认知和行为的一种综合的心理和生理状态, 是对外界刺激所产生的心理认知和反射, 以及相应的生理反射和表达。情绪是人类生活的重要内容, 因此致力于对人类情绪进行计算机感知和表达进行研究和探索的情绪计算是人工智能和智能人机交

**所属课题:** 高等院校博士点专项基金 (20122302120070); 国家自然科学基金项目 (61203378); 深圳市基础研究计划 (JCYJ20120613152557576); 深圳市国际合作计划 (GJHZ20120613110641217)。

**作者简介:** 徐睿峰 (1973—), 男, 副教授, 硕士生导师, 主要研究方向为自然语言处理; 邹承天 (1986—), 男, 硕士研究生, 主要研究方向为自然语言处理; 郑燕珍 (1988—), 女, 硕士研究生, 主要研究方向为自然语言处理; 徐军 (1980—), 助理研究员, 主要研究方向为自然语言处理。

互等研究的重要内容。考虑到文本中蕴含着精细化的情绪描述，特别是包含了情绪的刺激激活原因、刺激认知结果、情绪状态迁移和情绪外在表达等多方面信息，近年来陆续出现了分析文本中个体情绪状态迁移、利用文本发现情绪的触发机制和认知结果，以及对个体接收到文本中包含的刺激可能产生的情绪进行预测等方面的研究。这些研究显著提高了情绪计算研究的深度和广度，使其不止限于计算科学领域，更为心理学、认知科学和社会学等研究提供新的动力，因此，文本情绪计算正在成为新的研究热点。

情绪词典是文本情绪计算的基础资源。典型的工作包含英文中的 WordNet-Affect<sup>[1]</sup>以及中文的情绪词汇本体库<sup>[2]</sup>。对应于六种基本情绪 (anger/fear/joy/sadness/disgust/surprise)，WordNet-Affect 在 WordNet 中标注了对应词语和同义词集。中文情绪词汇本体库<sup>[2]</sup>，则将情绪分为七个基本大类和二十一个小类，对词语表达的情绪类别和强度分别进行标注。这些资源有力地推动了文本情绪计算的研究，但在应用中也暴露出一些问题。首先，这些词典通常对描述个体情绪和引发个体情绪的词语未做区分，导致某些词语在作为主体情绪表达输出或者作为主体接收到刺激产生的情绪之间的差异无法一致地记录在词典中。例如“鄙视”作为主体输出动作时，包含的主要情绪是“厌恶”；而作为主体接受到输入刺激时，激发的主要情绪是“愤怒”。如果不对词语作为主体的输出表达或者输入刺激进行区分，显然容易导致标注知识的疏漏或者冲突。其次，这些词典将词语视为一个整体标注，而由于词语语义歧义的存在，往往导致标注结果的歧义。例如“兴奋”可以表达人的高兴和激动的情绪，也可以描述一种无情绪的神经生理活动。可见现有基于词语的标注方案存在着缺陷。

为此，在对个体情绪产生、迁移机制及对应的文字表达进行分析的基础上，本文建立了基于“刺激认知-反射表达”的文本情绪计算框架，并在此框架下对情绪词语的功能和特性进行分析，探索了一种新型情绪词典建设方法。首先，针对词语歧义的问题，引入 HowNet 的词性和义项信息，将词语转换为按词性和语义区分的多个词条进行标注，以降低歧义。其次，对词语作为情绪表达和情绪认知的不同角度进行区分，对每个情绪词条进行基于角色的多维度情绪类别和强度标注。此外，对词语的不同情绪表达和情绪认知类型进行了细化分类和标注。目前已完成近 1300 个词语的标注工作，通过对词语已标注信息的观察显示，本文提出情绪表达和情绪认知分离的情绪词典建设体系可以对词语的情绪相关知识进行多角度的描述和表达，可以较好地支持多种文本情绪计算任务。此外，标注工作显示利用 HowNet 的词条进行标注可以显著提高标注结果的一致性。

## 2 相关研究

情绪词典是文本情绪计算的基本资源。目前的情绪词典的构建往往结合人工标注、语义词典扩展和基于语料库抽取标注方法。英语中最重要的情绪词典资源 WordNet-Affect 是通过选择和标注代表情绪概念的 WordNet 中的同义词集而获得的 Ekman 六种基本情绪相关的词语，然后利用 WordNet 中定义的关系、情绪标签和领域标签相互关系对其进行扩展，找到情绪同义词所在的 Synset 并扩展得到情绪词典。基于 WordNet-Affect，陆续有其它研究试图将其扩展到其它语言如 Bobicev et al. 通过机器翻译得到了罗马尼亚语和俄语的并行情绪词典<sup>[3]</sup>；Torii et al. 直接根据 WordNet-Affect 的同义词 ID 和对应的日语 WordNet 创建了日语情绪词典<sup>[4]</sup>；Xu et al. 利用机器翻译和基于对应翻译结果双语图过滤方法创建了中文种子情绪词典，再根据词语相似度计算从同义词词林中扩展该词典，获得词汇量较大、质量较高的中文情绪词典<sup>[5]</sup>。WordNet Affect 的缺陷在于仅对词语情绪类别进行标注，收录的词语多数为描述性情绪词，从它翻译和扩展而来的词典也并未作更多的表达性类别扩展，因此是不全面的情绪词典。

徐琳宏、林鸿飞等构建了中文情绪词汇本体库<sup>[2]</sup>，将情绪分为七个基本大类和二十一个小类。利用相关情绪词典和语义知识库获得候选情绪词，再人工对部分种子词语的情绪类别

和强度进行标注,最终使用基于互信息方法在大规模语料中对所有情绪候选词进行自动情绪标注。该情感词汇本体共 27467 个词语,每个词包括了词性、情绪类别、情绪强度值、情感倾向性等信息,该词典收录了较完整的表达性和描述性情绪词,但只对情绪词进行粗分类,同时对情绪表达和情绪认知未做明显区分。

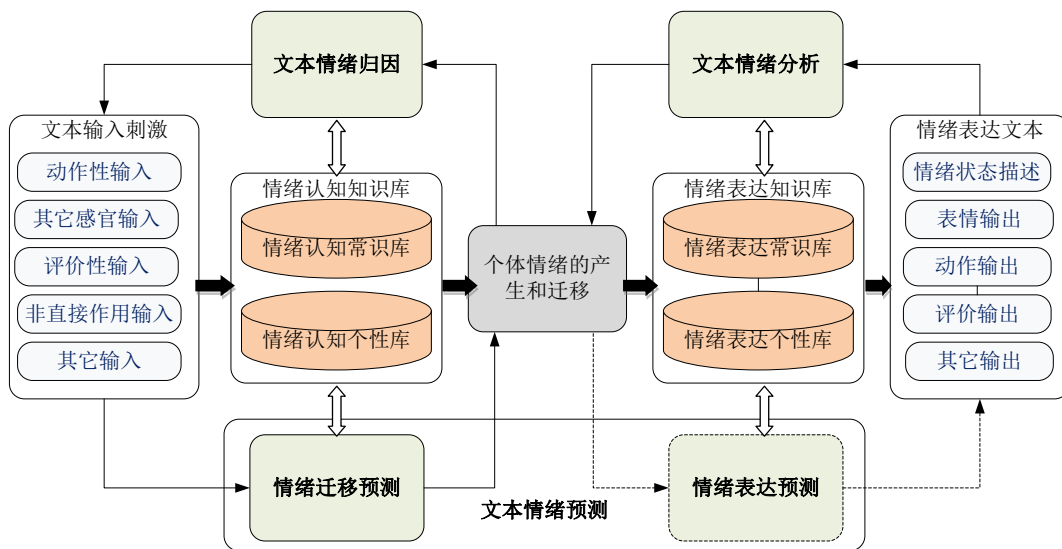
另外一类特殊的情绪词典是情绪认知词典,也就是记录可能诱发个体情绪的词汇。这些词汇可以包含或不包含特定情绪,但会暗示或者诱发某些情绪的产生。Svitlana Volkova 建立颜色情绪词典 CLex<sup>[6]</sup>,标注了每一种颜色诱发的情绪、以及潜意识里联想的某一种事物。Saif Mohammad 则对情绪诱发词语的歧义进行了研究<sup>[7]</sup>。

### 3 情绪词典标注体系设计

情绪词典资源是情绪计算的基础,而词典标注体系又是由情绪计算框架所决定的,因此要在情绪计算相关研究上有更大的突破首先需要建立合理的情绪计算框架,并在对应的资源建设上获得可靠支撑。

#### 3.1 基于“刺激认知-反射输出”机制的文本情绪计算框架

情绪心理学和认知科学的研究成果揭示出个体情绪的产生和迁移是由个体对接受到的动作性输入、其他感官输入、评价性输入等直接刺激和非直接作用刺激的认知结果决定的,而认知结果则是对输入刺激的常识性认知以及与个体情绪特质相关的个性化认知共同作用的结果。个体情绪的产生和迁移往往通过表情、动作、评价和其它方式进行反射和表达。“刺激激发认知、认知推动情绪产生和迁移、情绪迁移通过反射输出表达”构成了情绪产生、迁移和表达的基本机制。对应于这个机制,结合文本情绪计算的特点,本文提出了一种基于“刺激认知-反射输出”机制的文本情绪计算框架,如图 1 所示。



在此框架下,文本情绪分析研究从文本中识别个体情绪状态或者输出的描述,通过结合情绪表达常识及个体情绪迁移规律评估个体情绪状态的机制及对应计算方法;文本情绪归因研究利用个体情绪状态迁移为线索,从文本中发现和评估引起情绪迁移的刺激和认知的机制及对应计算方法;文本情绪预测研究从文本中识别可能引起个体情绪迁移的输入,利用对输入的情绪认知结果评估个体情绪迁移的趋势和情况的机制及对应计算方法。显然,对应于情绪表达和情绪认知,应分别建立对应的情绪词典。

#### 3.2 标注词语的选取

### 3.2.1 初始情绪词表

情绪词列表是通过合并多个已有情绪词典获得，主要包括：中文情感词汇本体库<sup>[2]</sup>、Xu et al. 从 WordNet-Affect 翻译并从同义词词林中扩展而来的情绪词典<sup>[5]</sup>、北京大学 Xu et al. 构建的 ICL 基本情绪词典<sup>[8]</sup>等获得。对于部分词语不符合本文定义的情绪词典覆盖范围，在标注过程这些词语会人工删除。

### 3.2.2 利用 HowNet 信息的词条生成

由于一些词语存在多个词性或者多个词义，而不同词性、词义表达的情绪可能完全不同，所以凭借词语本身无法确定标注信息。这里，引入 HowNet 提供的词性和义项信息对词语本身进行划分，解决一词多义的问题。例如“严峻”一词，具有相同词性的两个不同的含义，一个是用于形容形势的，而另一个是形容表情的，因此需要标注不同的情绪类别。此外，HowNet 中对词汇的英语翻译、汉语例句、以及对评价词和情感词倾向性的特殊说明也作为标注的重要参考。

## 3.3 情绪标注划分

### 3.3.1 情绪表达、情绪认知的基本划分方法

本研究将情绪词划分为情绪表达和情绪认知两个大类。情绪表达是指该词用于描述个体持有的某种情绪，或者对持有情绪的描述或者由持有情绪而产生的行为。而情绪认知指的是该词的作用对象会因为这一词汇所描述内容的刺激的认知结果而产生某种情绪。

1) 情绪表达词又可细分如下五种情况：

1. 个体自身情绪描述，如“悲伤”、“快乐”、“同情”。
2. 个体表情性输出，如“微笑”、“大哭”。
3. 个体动作性输出，如“破口大骂”、“怒吼”。
4. 个体评价性输出，如“混蛋”、“恶贯满盈”等。
5. 特殊个体评价性输出，如“感人”、“美丽”等。

2) 情绪认知又可细分为如下四种情况：

6. 个体接受到的动作性输入，如“殴打”“辱骂”“赞扬”等。
7. 个体接受到的非直接作用输入，如“载歌载舞”、“行凶”、“春晖”等。
8. 个体接受到评价性输入，如“混蛋”、“垃圾”等。
9. 个体自身经历的事件或者承受的状态，如“惨祸”、“失恋”等。

具体标注工作围绕这 9 个类别情绪属性进行的，即如何将情绪词按照这 9 个类别标注。

### 3.3.2 情绪词的角色属性

在进一步的研究与试标注的基础中，本文提出将上述八种情绪类型划分按照情绪主体角色属性的不同归纳为以下 4 类角色类别，分别用 ABCD 标记。

A. 发出评价者的情绪：通常是作者或说话人的情绪，包括情绪表达中的“4. 个体评价性输出”和“5. 特殊个体评价性输出”两类。这两种情况的共同特点是，作者或说话人对其他事物进行了含有主观感情色彩的评价。4 类含有的情绪较为明显，可以是说话人的感情宣泄，例如对于“混蛋”一词，反映了作者的厌恶和愤怒的情绪。而 5 类别的情绪词所含有的主观感情色彩相对较弱，通常作为客观描述，加上程度副词后带有明确的感情色彩。例如“这是个感人的故事”作为客观描述是无情绪的，而“真是太感人了”作为主观情绪表达带有明显情绪，类似的词语还有“悲惨”、“富丽堂皇”等。

B. 所属者或被描述者的情绪：包括情绪表达中的“1. 个体自身情绪描述”、“2. 个体表情性输出”和“3. 个体动作性输出”三个类别。这种情况的共同特点是，情绪词是情绪主体所处的状态或发出的动作，对词汇进行标注的情绪内容是该词汇所属主体的情绪。例如“他悲痛欲绝”“这个人怒吼道…”。

C. 动作、评价、事件、状态受体的情绪：包括情绪认知中的“6. 个人接受到的动作性输

入”、“8. 个体接受到评价性输入”和“9. 个体自身经历的事件或者承受的状态”三个类别。6与8的共同点是情绪词本身是一种由他人发出的动作或作出的评价；9与6的区别在于9是强调自身经历，而6是强调由外界施加并且是个体直接接受的动作。以下为典型例词，“殴打”，被殴打的人会感到悲伤和愤怒，被他人评价为“混蛋”通常会产生愤怒等负面情绪，经历“失恋”通常会产生悲伤、失望、烦闷等情绪，处于“形单影只”状态也会有悲伤等负面情绪。

D. 旁观者或读者的情绪：是指事件、动作、状态等的旁观者或者读者所产生的情绪，包括“7. 个体接受到的非直接作用输入”。这一类情况的情绪往往不是直接刺激作用于情绪主体而产生的，而是由于感知到的某种情况、状态产生了情绪。文中旁观者或者文章读者结合自己的常识、立场等产生相应的情绪，例如，“敲诈勒索”一词通常会令人产生厌恶和愤怒等情绪。

在情绪词典标注过程中，对每一个情绪词首先要进行其在不同情绪主体角色下所属的9个情绪属性的标注。标注的内容即为3.3.1描述的9个情绪属性的具体标号。通常情况下，一个词会兼有以上四种情绪主体角色的情绪类别中的一种或几种，例如，“混蛋”一词的类别标注为4/0/8/7。其含义是，“混蛋”一词的评价发出者、评价接受者以及旁观者或读者都会有某种情绪。其中0代表没有或无法确定这一类别下的情绪，实际标注中通常略去。

上述归纳划分方法在一定程度上解决了表达/认知类别标注混乱的问题，有利于标注者清晰思路和理顺逻辑，从而提高多人标注的一致性，更重要的是，它能为实际应用该词典解决常见情绪研究问题提供可靠的依据。

### 3.3.3 情绪词标注方案

汉语里一个词往往有多种不同含义和用法，仅仅通过词性以及其它HowNet给出的信息仍不足以严格的判断一个情绪词需要进行哪些类别情绪内容的标注。例如，作为A类的用于评价的词，可能是形容词，如“他干得**不赖**啊”，也可能是名词，如“XX是个**混蛋**”，甚至可能是动词，如“没完没了地**扯皮**，把事情都耽误了”。由此可见，需要分别针对各类词性的情绪词制定不同的类别划分方法。本标注方法在这一问题上，总的原则是尽可能的对词的各种用法进行全面的考量，只要该词有可能被用于符合某个角色类别的情况，就对其在此类别下的具体情绪属性类别及具体情绪内容进行标注。表2是各个词性类别对应情绪属性的例词，如下所示：

表2 词性对应各情绪属性的例词

角色属性类别	情绪描述					情绪认知			
	A		B			C			D
情绪属性/词性	4	5	1	2	3	6	8	9	7
名词	混蛋、大赞	慈爱、无辜	希望、悲伤、痛苦	眼泪、笑声	怒斥、溢美之词	谩骂、表扬、耳光	混蛋、垃圾	惨祸、挫折、纠纷	死难者、车祸、兽行
动词	扯皮	牺牲、与世长辞	喜欢高兴、失望	瞠目结舌、大笑	怒吼、喜获、赞赏	羞辱、殴打、赞赏	扯皮、坑蒙拐骗	含冤、束手无策	行凶、载歌载舞
形容词和副词	不赖、超赞	感人、富丽堂皇	悲伤、萎靡不振				无耻、卑鄙、下贱	孤零零、美满、悲惨	惨绝人寰、满目疮痍

### 3.3.4 具体情绪类别及其强度

在确定情绪词所属的类别信息之后还要对其包含的情绪信息的具体内容进行标注。包括两方面的内容：情绪类别及其强度。情绪类别指的是通常所说的基本情绪，如快乐、悲伤、愤怒、厌恶等，具体类别划分与大连理工大学的情感词汇本体<sup>[2]</sup>中使用的体系相一致，共7大类21小类，包括：乐（快乐、安心），好（尊敬、赞扬、相信、喜爱、祝愿），怒（愤怒），哀（悲伤、失望、疚、思），惧（慌、恐惧、羞），恶（烦闷、憎恶、贬责、妒忌、怀疑），

惊（惊奇）。

在实际标注过程中发现，只有相对少量的描述基本情绪的词具有单一的情绪类别，大部分的词语其内涵情绪都是多种情绪的搭配组合，常见的如“赞扬”、“喜爱”与“尊敬”、“憎恶”与“贬责”等。对于具体每个情绪类别的强度，本文采用 0, 1, 3, 5, 7, 10 的数值标注，用数值的大小表示情绪的强烈程度，0 表示没有该情绪，5 表示有明显的此种情绪，10 代表情绪的强度最为强烈。完成多人标注后，最终的强度值是由多人的标注数值求均值所得。

### 3.3.5 完整标注示例

综合上述标注体系，一些典型例词的实际标注效果如下所示：

表 3 标注实例

情绪词	词性	HN 义项	类别编号	快乐	赞扬	愤怒	悲伤	憎恶	贬责	惊	...
高兴	V	joyful 喜悦	0/1/0/0	0/5/0/0							
难过	ADJ	difficult 难:scope=GoT...	0/0/0/0								
难过	V	sorrowful 悲哀	0/1/0/0				0/5/0/0				
难过	V	unfortunate 不幸	0/3/0/0				0/5/0/0				
混蛋	N	human 人:modifier=des...	4/0/8/7			7/0/5/0		5/0/3/3	3/0/0/0		
怒吼	V	MakeSound 发声	0/3/0/0			0/7/0/0					
怒吼	V	cry 喊	0/3/0/0			0/7/0/0					
瞠目结舌	V	stupefied 木然	0/2/0/0							0/7/0/0	
羞辱	V	damage 损害	5/0/6/0			3/0/3/0	0/0/7/0	3/0/5/0	5/0/0/0		
羞辱	N	emotion 情感:CoEvent=shy	0/1/0/0			0/3/0/0	0/7/0/0	0/5/0/0			
敲诈勒索	V	cheat 骗	0/6/7/0			0/5/3/0					
死难者	N	human 人:{suffer 遭受:	0/0/0/7				0/0/0/5				
满目疮痍	ADJ	miserable 悲惨	4/0/0/7				5/0/0/5	5/0/0/3	5/0/0/3		
不赖	ADJ	HighQuality 优质	4/0/8/7	3/0/5/3	5/0/0/3						
孤零零	ADJ	{lonely 孤}	5/0/9/7				3/0/5/3				
束手无策	V	{BeUnable 无能}	0/0/9/7				0/0/5/3				
纠纷	N	{fact 事情:CoEvent={... 争吵}}	0/0/9/7			0/0/5/0		0/0/5/3	0/0/0/5		

注：由于篇幅所限，上例省略了部分 HowNet 义项内容以及部分情绪类别，同时并未按照词义对近义情绪词进行归并。

### 3.3.6 词典应用方法

与本研究所构建的情绪词典相比，传统的情绪词典通常并不对情绪的角色属性加以区分，对各种类型的情绪词都予以同等的处理方式。而在实际应用中，针对不同的文本情绪研究任务，每个情绪词所体现出的作用也是不同的。从目前已有情绪研究所涉及的不同角度来看，可以分为以下三大类：

- 1) 针对传统的作者情绪分析任务：发挥作用的情绪词主要体现在情绪角色属性为 A 所标注的内容；
- 2) 针对文本细粒度情绪分析任务：在分析文中角色的情绪内容时主要采用情绪词的角色属性类别为 B、C、D；
- 3) 针对读者情绪预测任务：需要重点关注的是情绪词在 D 类别下标注的情绪内容。

例如“敲诈勒索”一词，如果需要分析的是文中所描述的“遭到敲诈勒索”的人的情绪，则关注的是该词在 C 类别下所标注的情绪（在此为情绪属性类别 6），即愤怒、恐惧等；而对于读者情绪预测任务则需要重点关注 D 类别下（即为情绪属性类别 7）标注的内容，即贬责、憎恶、愤怒等。

#### 4. 标注结果统计与分析

第一阶段标注选取了较典型和明确的情绪词进行标注，借助 HowNet 中正面情感词、负面情感词的标注信息进行挑选，总计 1259 个词语，对应获得 1500 个 HowNet 义项，以下表格为各项标注结果的统计数据。

表 4 是对情绪属性、情绪类别对应的词语数及平均情绪强度值统计，总体上主要为表示情绪描述的词语，而情绪认知为应激情绪词使用相对较少，符合常规语言现象。标注的词语中情绪类别主要为：快乐（323）、喜爱（351）、愤怒（310）、悲伤（566）、烦闷（295）、憎恶（394）、贬责（342），这些是代表了具有明显褒贬倾向的情绪类别，也是在文章中较常表达的情绪。对于主要的情绪属性和情绪类别来说平均情绪强度值分布在 3.5~4.5，说明大部分情绪表达均在中等强度水平，太过强烈或者轻微的情绪表达较少。

表 4 情绪属性与情绪类别标注分布统计结果（词语数及平均情绪强度值）

角色类别	情绪描述					情绪认知				总计
	A		B			C			D	
情绪类别	4	5	1	2	3	6	8	9	7	
快乐	16(4.44)	11(2.73)	107(5.04)	16(6)	43(4.77)	44(4.23)	5(3.4)	27(4.89)	54(3.81)	323(4.59)
安心	2(2)	43(2.95)	62(4.58)	5(2)	12(3)	40(3.7)	14(3.21)	27(3.89)	61(3.52)	266(3.66)
尊敬	6(3.5)	17(3.12)	21(5.57)	0	4(5.25)	1(2)	0	0	28(3.29)	77(3.97)
赞扬	38(4.71)	88(4.09)	6(4)	0	29(6.1)	2(2)	0	1(3)	112(3.36)	276(4.07)
相信	1(2)	7(2.71)	27(2.63)	0	2(3)	0	2(2)	2(2)	9(2.33)	50(2.54)
喜爱	30(3.93)	53(3.21)	130(4.83)	3(4)	30(4.03)	2(2.5)	0	15(2.8)	88(3.27)	351(3.94)
祝愿	1(4)	11(2.64)	4(3.25)	0	11(3.27)	2(3)	0	0	128(3.17)	157(3.15)
愤怒	23(2.78)	1(2)	64(5.02)	8(6)	49(5.41)	83(4.31)	13(3.15)	4(5.25)	65(3.69)	310(4.39)
悲伤	11(4.36)	34(3.59)	109(4.81)	6(4.67)	40(6)	53(3.28)	0	72(5.44)	241(3.62)	566(4.24)
失望	5(3.2)	3(3)	44(4.73)	1(6)	10(3.7)	8(2)	0	16(4.06)	17(2.88)	104(3.9)
疚	2(6)	0	22(4.55)	0	4(5.25)	0	1(2)	0	0	29(4.66)
思	0	3(2.67)	25(3.36)	0	6(2.17)	1(2)	0	3(2)	5(2.8)	43(2.95)
慌	2(2)	8(2.63)	46(4.11)	0	1(2)	20(3.3)	0	7(2)	41(2.24)	125(3.1)
恐惧	3(5)	7(3.86)	46(5.22)	0	0	23(3.83)	0	7(4.71)	28(3)	114(4.27)
羞	0	0	20(4.25)	5(5)	0	12(2.92)	11(3.45)	0	0	48(3.81)
烦闷	5(3.6)	7(2)	159(3.93)	6(4.83)	15(3.67)	39(2.9)	7(2)	48(3.94)	9(2.89)	295(3.67)
憎恶	40(4.55)	12(2.92)	50(4.42)	0	42(3.86)	82(3.54)	13(4)	6(5.33)	149(3.91)	394(3.95)
贬责	44(5.34)	30(3.87)	8(2.38)	0	33(3.45)	25(2.52)	0	5(4)	197(4.18)	342(4.07)
妒忌	0	0	5(3.6)	0	0	0	0	0	0	5(3.6)
怀疑	1(2)	5(3.8)	13(3.38)	0	1(2)	0	0	0	7(2.57)	27(3.15)
惊奇	16(4.13)	30(4.4)	24(5.33)	5(7.6)	2(5.5)	2(4)	0	1(7)	47(3.51)	127(4.37)
总计	246(4.31)	370(3.49)	992(4.52)	55(5.31)	334(4.56)	439(3.56)	66(3.23)	241(4.42)	1286(3.56)	

表 5 是对情绪属性与词性进行相关性统计，结果显示作为情绪属性 4 和 5 的情绪词大部分为形容词，符合评价词多为形容词的特性；作为情绪属性 1、2、3 的情绪词大部分为动词，符合情绪所属者表达自身情绪多为心理、表情、动作的特性，而这些词多为动词；作为情绪属性 6、8 的情绪词大部分为动词和形容词，与 3 和 4 的词性相对应；作为情绪属性 7 的情

绪词多为动词、形容词，符合读者或旁观者由于动作、评价输出引起相应情绪的特性；作为情绪属性 9 的情绪词多为动词、名词、形容词，符合表示事件、状态等词语的词性分布。

表 5 情绪属性与词性标注分布统计结果（词语数）

角色类别	情绪描述					情绪认知				总计
	A		B			C			D	
情绪属性/词性	4	5	1	2	3	6	8	9	7	
名词	9	13	53	14	0	6	0	117	592	804
动词	35	37	857	37	332	421	0	61	199	1979
形容词	189	310	80	4	1	10	66	61	469	1190
副词	7	3	8	0	0	0	0	0	1	19
其他	6	2	0	0	1	2	0	0	6	17

表 6 是对词性与情绪属性兼类标注分布的统计结果。由于中文词语表达的多样性和复杂性，情绪属性兼类的情况也比较多，各类组合兼类共有 19 种，现阶段已标注的典型情绪词中主要为 1. 个体自身情绪描述的词语，兼类主要分布在 3/7、4/7、4/7/8、5/7、6/7、7/9，并且大部分兼类中都包括属于情绪认知的情绪属性 7，代表对读者/旁观者情绪的考虑，而这类标注内容将在情绪预测和基于角色的情绪分析中得到充分的应用和验证，体现了本词典的新颖特性。

表 6 词性与情绪属性兼类标注分布统计（词语数）

兼类	名词	动词	形容词	副词	其他	总计	兼类	名词	动词	形容词	副词	其他	总计
1	26	476	27	1		530	4/6/7		4	1			5
1/4/7		3	3			6	4/7		2	52	3	2	59
1/5/7		6	13			19	4/7/8			22			22
2	9	31	3			43	5			23		1	24
3		39	1			40	5/6/7		1				1
3/4/7		1				1	5/7	6	12	119	5		142
3/5/6/7		1				1	5/7/8			10			10
3/5/7		2				2	5/7/9			12			12
3/6		58				58	6	1	8			1	10
3/6/7		13				13	6/7	2	75	3			80
3/7		63			1	64	7	61	74	10		1	146
4	1	1			2	4	7/9	37	12	16		1	66
4/6		2				2	9	5	2	1			8
9/7	7	10	2			19	总计	155	896	318	9	9	1387

## 5. 总结与展望

目前情绪词典的建设已完成标注方案设计和完善，通过多轮试标注，对标注体系不断完善。先后解决了一词多义带来的歧义性、近义词分散带来标注的不一致性、角色观测角度不同带来的不确定性等问题，并且对每个情绪词尽可能全面地考量各种用法。目前正在第一阶段标注的基础上进行进一步完善，并逐步完成全部情绪词典标注。



相比现有情绪词典,本文提出的情绪表达和情绪认知分离的情绪词典建设方案可以对词条的不同角度观测到的情绪状态和强度进行区分,显著降低了传统方法由于描述角度不一致带来的问题。同时,基于 HowNet 义项的标注也降低了由于词义歧义带来的情绪标注歧义。此外,这种基于义项的标注提高了标注结果的精度和一致性。本文开发的新体系情绪词典可以更好地支持文本情绪计算研究。后期工作将逐渐扩展到更多词汇进行标注,增加标注工作参与人数,利用对标注结果进行多人综合的方法,增强标注可信度。同时计划对大量词汇进行半自动标注,以提高标注效率。

## 参考文献:

- [1] C. Strapparava, and A. Valitutti. WordNet-Affect: An Affective Extension of WordNet [C]// Proceedings of LREC 2004, 1103-1106
- [2] 徐琳宏, 林鸿飞, 潘宇, 任惠, 陈建美. 情感词汇本体的构造[J]. 情报学报, 2008, 27(2):180-185
- [3] M. Sokolova, V. Bobicev. Classification of Emotion Words in Russian and Romanian Languages [C]//Proceedings of RANLP-2009, 415-419
- [4] Y. Torii, Dipankar D., S. Bandyopadhyay, and M. Okumura. Developing Japanese WordNet Affect for Analyzing Emotions [C]// Proceedings of ACL-WASSA 2011, 80-86
- [5] J. Xu, R.F. Xu, et al. Chinese Emotion Lexicon Developing via Multi-lingual Lexical Resources Integration [C]// Proceedings of CICLing 2013, 174-182
- [6] Volkova S, Dolan W B, Wilson T. CLex: A Lexicon for Exploring Color, Concept and Emotion Associations in Language [C]// Proceedings of EACL 2012, 306-314
- [7] S. M. Mohammad, P. D. Turney .Emotions Evoked by Common Words and Phrases: Using Mechanical Turk to Create an Emotion Lexicon [C]// Proceedings of NAACL-HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text, 26-34, 2010
- [8] G. Xu, X. F. Meng, H. F. Wang. Build Chinese Emotion Lexicons-Using A Graph-based Algorithm and Multiple Resources [C]// Proceedings of COLING 2010, 1209-1217