

面向情感分析的构式主观态度义初探

黄思思, 詹卫东

(北京大学中文系/中国语言学研究/计算语言学教育部重点实验室 北京 100871)

摘要: 文本情感分析是目前自然语言处理领域的前沿研究课题。本文考察了两类跟情感表达有关的构式 [+程度]构式(42条)和 [+主观态度]构式(185条),前者表达情感的强度;后者的表达功能具体包括三种情况:评价、情感和立场。评价是指依据一定的标准对某实体或事件做出价值判断。情感是指对外界刺激肯定或否定的心理反应,如喜欢、愤怒、悲伤、恐惧、爱慕、厌恶等。立场是指对某人的观点或意愿表达接受、反对或不置可否。一个具体的 [+主观态度]构式可以只包含上述一种情况(如只表达评价),也可以包含不止一种情况(如同时表达情感和立场)。有的 [+主观态度]构式除了上述主要表达功能之外,还有附加语义特征(如同时表达情感和后果)。为了帮助计算机做文本情感分析,本文设立了主观态度语义要素7元组 <主体,对象,类型,强度,动因,后果,命令>,并讨论了从句子中抽取这些语义要素的规则。

关键词: 情感分析; [+程度]构式; [+主观态度]构式; 句法形式特征

中图分类号: H030 文献标识码: A 文章编号: 1000-5544(2018)06-0027-07

Abstract: Sentiment analysis is a frontier research topic in the field of natural language processing. This article discusses two types of constructions related to sentiment analysis: degree constructions and sentiment constructions. We gathered a grand total of 227 instances of both types, including 42 degree constructions and 185 sentiment constructions. While degree constructions can alter sentiment intensity, sentiment constructions can convey sentiment independently. The latter contains three subtypes: evaluation, emotion and standpoint. Evaluation refers to judging the value of an entity or event based on certain criteria. Emotion refers to positive or negative psychological reactions to external stimuli, such as like, anger, sadness, fear, love, disgust and so on. Standpoint refers to one's views or wishes, containing support, opposition, or be noncommittal. A specific sentiment construction may contain only one of the above semantic features, e. g., expressing evaluation only, or more features, e. g., expressing emotion as well as standpoint. Some sentiment constructions can contain more semantic features, e. g., expressing emotions as well as consequences in addition to the main expressive functions described above. In order to help the computer to do the task of sentiment analysis, this paper also discusses the semantic elements of sentiment <subject, object, type, strength, motivation, consequences, command> and the syntactic realization of each element.

Key words: sentiment analysis; degree construction; sentiment construction; syntactic feature

DOI:10.16362/j.cnki.cn61-1023/h.2018.06.005

1. 引言

情感分析(sentiment analysis)是当今自然语言信息处理领域的热门课题。一般来说,广义的情感分析包括分析文本中包含的说话人的态度,而狭义的情感分析则主要指的是针对说话人关于某物或某事的观点进行“赞同”或是“反对”的分类(周立柱等 2008: 2725; 孙毅 2018: 34)。本文采用广义的情感分析的概念。

情感词在情感分析中至关重要,但是仅仅依靠情感词进行情感分析是远远不够的。在真实文本中程度副词、否定副词等修饰词会加强、减弱或置反情感词的极性,因而学界开始探索短语的情感分析。短语的情感有的可以由规则推导出来,如例(1)^①。但是,一些短语的情感无法通过规则推导出来,如例(2)、例(3)。

(1) 她很漂亮。

(2) 她要多漂亮有多漂亮。

(3) 好你个李云龙!

如果将例(2)、例(3)当做常规短语结构处理,会导致情感信息分析不全甚至错误。例(2)不包含程度词却表示“漂亮到极点”。如果当做常规短语处理会丢失“情感强度”信息。例(3)包含正面评价词“好”,但是整个结构却往往表达负面评价。如果当作常规短语处理会导致情感信息分析不准确。侯敏等(2013: 107)在研究评价短语时曾提及将上述单位直接收入评价短语词典。现在的一些情感词库,例如《<知网>情感分析用词语集》^②、大连理工大学的情感词汇本体库^③等,也搜集了少量短语级别的单位。但是,现代汉语中这种单位的大致数量、情感类型、在情感分析中的作用尚缺乏系统研

究。本文希望结合构建现代汉语构式知识库^④的工作,在这方面做进一步的探索。Goldberg(1995:4)对“构式”所下的定义是“C是一个构式,当且仅当C是一个形式和意义的匹配体<Fi, Si>,而其形式Fi也好,意义Si也好,所具有的某些特征不能全然从C的组成成分或先前已有的其他构式所推知。”有的构式对情感分析没有贡献,如“v+来+v+去”(看来看去)。该构式具有“反复”义,“看来看去”就是“反复看”的意思。“反复”义是对动作量的客观描述,本身不表达情感,也不能改变情感的强度。这种构式,本文不予考虑。本文主要考察跟情感表达有关的构式。下面第2节概述跟情感表达有关的构式及相关意义要素分析流程,第3节分析主观态度的类型,第4节讨论情感的程度分级,第5节讨论主体、对象、动因、后果、命令等要素,第6节是结语。

2. 跟情感表达有关的构式及相关意义要素提取流程概述

本文讨论的跟情感表达有关的构式具体包括以下两种:①[+程度]构式,整个构式包含“程度”特征,与情感词搭配可以表达情感的强度。例如“要+多+a+有+多+a^⑤”(要多漂亮有多漂亮),整个构式的意义是“a到极点”。②[+主观态度]构式,整个构式独立表达评价、情感或立场。例如“好+你+个+np”(好你个李云龙),整个构式往往表达负面评价。

现有情感词典在描述情感词时,或者仅仅区分正面、负面;或者分为正面评价、负面评价、正面情感、负面情感四类(例如,《<知网>情感分析用词语集》);或者将情感细分类,例如,大连理工大学的情感词汇本体库将情感分为7大类21小类。与词不同,构式可以表达复杂态度,且不同构式表达的态度之间存在细微差异。用现有描述体系无法准确描述构式表达的主观态度。例如:

- (4) 他不来就不来。
(5) 他爱来不来。

例(4)、例(5)表达的主观态度有同有异。两者都表达对某人的行为不满、愤怒等负面情感。两者的差异在于立场,即是否接受/赞成^⑥对方行为。例(4)表示“接受/赞成”对方行为,例(5)表示对对方行为“不置可否”。按照现有描述体系,即仅仅从正面/负面或者仅仅从情感分类的角度出发,不便描述两个构式表义特征的差异。所以,在描述构式的主观态度时需要增加“立场”这一特征。另外,构式可以表达复杂态度,例如上述例(4)、(5)既包含情感又包含立场。不适合按照意义分类的方式进行描述,而应该采用特征的方式,即不是将构式分为评价构式、情感构式、立场构式,而是说某一[+主观态度]构式包含哪几个语义特征,以及每一个特征的取值是什么。例(4)、(5)两个构式的主观态度可以描述如下:

不v就不v 情感:不满、愤怒等,立场:接受
爱v不v 情感:不满、愤怒等,立场:不置可否

本文将评价、情感、立场三者统称为主观态度而没有沿用情感分析中将三者统称为情感的做法。这样处理一是为了避免混淆。如果用情感构式统称评价构式、情感构式、立场构式,就会导致类名跟成员名相同,会引起混淆。二是将评价、情感、立场合称为态度也符合社会心理学对态度的定义。侯玉波(2007:87-88)对态度的定义包含三个成分:1)认知成分:它指人们对外界对象的心理印象,包括有关的事实、知识和信念,认知成分是态度其余部分的基础。2)情感成分:它指人们对态度对象肯定或否定的评价以及由此而引发的情绪情感,情感成分是态度的核心与关键,情感既影响认知成分,也影响行为倾向成分。3)行为倾向成分:它指人们对态度对象所预备采取的反应,具有准备性质。行为倾向成分会影响人们将来对态度对象的反应,但它不等于外显行为。它们之间的关系可用图1表示(虚框中的为态度,态度不等于外显行为)^⑦。其中“评价”对应图1中的“观点、直觉、评价”等“情感”对应图1中的“情感、情绪”等“立场”对应图1中的“意向、倾向、偏好”等。

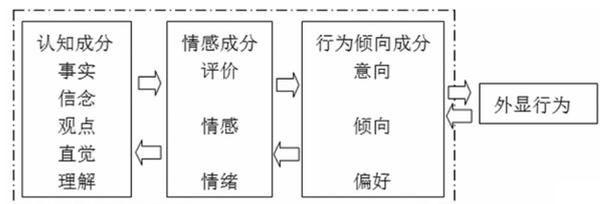


图1. 态度的成分及其关系

现有计算机情感分析任务中一般将情感要素表示为4元组<主体,对象,类型,强度>。考虑[+主观态度]构式的话,情感要素可以扩展为7元组<主体,对象,类型,强度,动因,后果,命令>。情感要素抽取有多种方法,基于句法分析的方法进行情感要素抽取大致有三个主要模块:1)分词、词性标注等预处理;2)构式识别以及句法语义分析;3)抽取情感要素。情感要素抽取的大致流程见图2。步骤①中分词、词性标注等技术已经比较成熟,这里不再赘述。步骤②需要借助经过改进的句法分析器。如何改进现有句法分析器使其能对构式进行句法分析需要另辟专文讨论。初步工作可以参见詹卫东(2017:233-234)对构式形式的相关论述。步骤③在步骤①、②的基础上完成情感要素的抽取。其中类型、强度两要素的取值主要取决于具体[+主观态度]构式以及[+程度]构式的特征描述,这部分信息存储在构式知识库中。其余要素的抽取,需要借助构式所在句子的主语、构式常项、变项与情感要素之间的对应关系^⑧(详见第5节)。

按照图2所示的流程,对“他爱来不来”进行分析,可以从该句中抽取对象、类型、动因等3个要素,其余几个要素的值无法确定,标记为“无”。

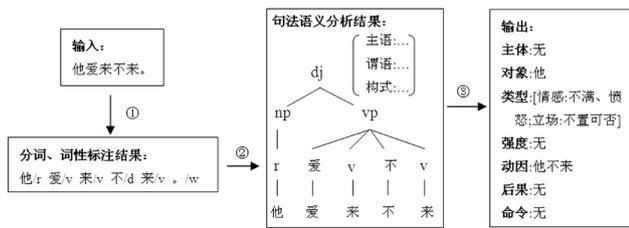


图 2. 情感要素抽取流程示意图

3. 主观态度类型的分析

主观态度包括评价、情感、立场三种。Thompson & Hunston(2000: 5)认为评价是“作者/说话者对命题或实体的感情、观点、态度或态势的表达”。就Thompson & Hunston(2000: 5)对评价的定义而言,评价的对象包括命题和客体,评价的内容包括感情、观点、态度等。这是一种广义的理解。我们采用评价的狭义的理解。评价,是人把握客体对人的意义、价值的一种观念性活动(冯平 1995: 1; 邵春 2016: 67-75)。通俗来讲,评价是指对某实体或事件依据一定的标准作出价值判断的过程。评价包括评价主体、评价对象、评价标准、价值判断等要素。在交际中,评价主体对应说话人,评价对象往往是句子主语,评价标准一般不出现,价值判断分为正面评价、负面评价两种。例如:

(6) 小王说“小李这个人很够朋友。”

(7) 小王说“小李这个人不怎么样。”

在例(6)、例(7)中“小王”是说话人,“小李”是评价构式描述的对象,因此“小王”是评价主体,“小李”是评价对象,评价标准隐含不出现。例(6)是正面评价,例(7)是负面评价。

情感是指对外界刺激肯定或否定的心理反应,包括喜欢、愤怒、悲伤、恐惧、爱慕、厌恶等具体的情感类型^⑨(《现代汉语词典》第5版: 1116)。情感意义可进一步分析为情感主体、情感对象、情感动因、情感类型等要素。在交际中,情感主体对应说话人,情感对象和情感动因可能出现也可能不出现,情感类型由构式的整体语义决定。例如:

(8) 小王说“我的天哪!”

(9) 小王说“小李也真是的,这么晚还不来。”

在例(8)、例(9)中,情感主体都是“小王”,情感动因都未出现。在例(8)中,情感对象没有出现,在例(9)中,情感对象是“小李”,即“小王对小李不满”。例(8)表达“吃惊”,例(9)表达“不满、愤怒”。在我们搜集到的[+情感]构式中表达“不满、愤怒”的构式占大多数。

本文中的“立场”是指对某人的观点或意愿表达接受、反对或不置可否。《现代汉语词典》(第5版: 838)对“立场”的定义是:认识和处理问题时所处的地位和所抱的态度。立场的语义分析可以包括立场持有者、立场对象、立场类型等要素。在交际中,立场持有者对应说话

人,立场对象可能是观点或行为,立场类型取值有接受、反对和不置可否三个。

(10) 小王说“可不是嘛。”

(11) 二大开口了。他声音和平得象念经文。“葡萄,你睡你的去,啥事不愁。要愁早该愁了。最愁人的都过去了。”(严歌苓《第九个寡妇》)

例(10)、例(11)中的立场持有者都是“小王”。在例(10)中,“小王”同意对方的观点。在例(11)中,葡萄想去睡而心怀疑虑,“二大”说出构式“你睡你的去”^⑩用以打消其疑虑,即“二大”赞成其意愿。

[+评价]、[+情感]、[+立场]三个语义特征各自独立,上面的例(6)一例(11)都只包含其中一个语义特征。但是有些[主观态度]构式包含不止一个语义特征,下面举两种包含两个语义特征的构式的例子。

有的构式既包含[评价]特征又包含[情感]特征,例如:

(12) 小王说“这什么大学生。”

在例(12)中,构式“什么大学生”既包含对某人的评价(即认为“某人作为大学生不合格”)也表达了说话人“不满、愤怒”等负面情感。

有的构式既包含[情感]特征又包含[立场]特征,例如:

(13) 小王说“你来就来吧。”

(14) 小王说“你爱来不来。”

(15) 小王说“你来什么来!”

例(13)一例(15)中的构式都表达了说话人相同的情感(即“不满、愤怒”等负面情感),但是表达了不同的立场。例(13)表达的立场是接受,即接受“你来”;例(14)表达的立场是“不置可否”,即“我才不管你来不来呢”;例(15)表达的立场是拒绝,即“你不要来”。为了方便称说,我们称例(13)为容忍构式,例(14)为不管构式,例(15)为驳斥构式。

上述例(6)一例(15)包含“主观态度”中的一种或两种语义特征。有的构式还包含其他语义特征。例如:

(16) 小王说“你要不来,看我不打死你。”

(17) 小王说“吃你的饭!”

(18) 小王说“你给我出去!”

在例(16)中,构式“要X,看我不vp”通过说明听话人行为(即“不来”)的后果(即“打死你”)来表达说话人对听话人行为的负面情感以及拒绝立场。为了称说方便,我们将例(16)中的构式称为威胁构式。威胁构式包含[+情感][+立场][+后果]三个语义特征。威胁构式与驳斥构式表达了相同的情感(即“不满”“愤怒”等负面情感),表明了相同的立场(即“拒绝”),两者的不同之处在于威胁构式还说明了听话人行为的后果。

例(17)除了表达不满、愤怒等负面情感,表明立场(即“反对”听话人当前行为)外,还发出了命令(即“吃饭”)。例(17)中的构式包含[+情感][+立场][+命

令]三个语义特征。

例(18)表达了不满、愤怒等负面情感,没有表明立场,但是发出了命令(即“出去”)。例(18)中的构式包含[+情感][+命令]两个语义特征。例(17)、例(18)都表达负面情感并且都表达命令,为了称说方便可以统称为强令构式。

有一点需要注意,虽然威胁构式和强令构式都表达了命令,但是威胁构式是让听话人放弃当前行为(也是引起说话人不满的行为),而强令构式是让听话人做其它的事情。可以对比例(16)与例(18)来体会这种差异。在例(16)中说话人对“听话人不来”不满;但是在例(18)中,说话人并不是对“听话人出去”不满,而是说话人对听话人当前行为不满并让听话人“出去”。

所以,总结起来看,[+主观态度]构式可以分为以下三组:

- 1) [+评价]构式、[+情感]构式、[+立场]构式
- 2) [+评价][+情感]构式、[+情感][+立场]构式
- 3) [+情感][+命令]构式、[+情感][+立场][+命令]、[+情感][+立场][+后果]构式

第1)组构式只包含[主观态度]中的一种语义特征,第2)组构式包含[主观态度]中的两种语义特征,第3)组构式包含[主观态度]中的一种或两种语义特征,并附加了[命令]或者[后果]语义特征。

4. 情感程度量级的分析

在文本情感分析领域,对程度词的分级比较有代表性的是《<知网>情感分析用词语集》,将程度分为6级:极其/最(如“百分之百,倍加”)、很(如“不少,不胜”)、较(如“多,更”)、稍(如“点点滴滴,怪”)、欠(如“半点,不大”)、超(如“过度,过分”)。

在汉语本体研究领域,张国宪(1996:34-41)提出,程度词可以从情态角度分出来表客观量的程度词和表主观量的程度词两个次类。前者适宜于以某客观事物为参照的比较,可以出现在比较句中。常见的表客观量的程度词有:最、更加、比较、略微等。后者不能用于以客观事物为参照的比较句,只适宜于以说话人心中的标准(主观)为参照作比较,以此来把握量的规定性。常见的表主观量的程度词有:极、很、有点儿等。主观记量和客观记量都可以分为微量、中量、高量和极量四个量级。表微量的程度词包括“稍微、略微、有点儿”等,表中量的程度词有“比较、较为”等,表高量的程度词常见的有“更加、更为”等,表极量的程度词常见的有“最、顶、最为”等。

本文收集到42条具有程度表达功能的构式。考察这些构式的量级,没有发现表“微量”和“中量”的。[+程度]构式的量级主要集中在极量(用“极”代表)、高量(用“很”代表)两级,另外,否定范畴与高量程度构式结

合,可以表达否定高量,相当于《知网》程度分级中的“欠”级(用“不很”代表)。构式的量级情况可以总结为表1:

表1. 程度构式按照量级分类

量级	构式	数量
极其	a+得+不能+再+a(好得不能再好);a+到+不能+再+a(好到不能再好);要+多+a+有+多+a(要多漂亮有多漂亮);再+a+不过(再好不过);再+a+也+没有+np+a(再帅也没有他帅);见+过+vp+的,没+见+过+np+这么+vp+的(见过不要脸的,没见过他这么不要脸的);别+提+多+vp+了(别提多倒霉了);说+不+出+的+vp(说不出的忧伤);X+就+别+提+了(那高兴就别提了);X+就+不+用+说+了(那高兴就不用说了)	10
很	一百个+vp(一百个不放心);一肚子的+ap(一肚子的委屈);满脸的+ap(满脸的不高兴);那+个+vp+啊(那个帅啊);那+叫+一+个+a(那叫一个帅);像+np+一样+似的(像仙女一样/像仙女似的);np+一样+似的(仙女一样/仙女似的);np+vp(仙女下凡);赛+np(赛华佗);小+np(小华佗);np1+的+np2(中国的乔丹);np+第+二(乔丹第二);vp1+vp2+了(丑死了);vp1+vp2+np+了(气死我了);vp+得+X(好得很);vp+得+np+X(急得我直跺脚);把+np+vp+得(把我气得);vp+得+np(气得我);这+np+vp+得(这衣服脏得);vp+到+X(好到爆);比+np+还/更+np(比专家还/更专家);n+中+的+n(专家中的专家);你+说+a+不+a(你说倒霉不倒霉);a+着+呢;够+a(够黑);g1+了+g2+了(倒了霉了);a+而+又+a(外行而又外行);看+把+np+vp+得(看你乐得);看+np+vp+得(看你乐得)	29
不很	不+怎么+a(不怎么漂亮);a+不+到+哪里+去(好不到哪儿去);能+a+到+哪里+去(能好到哪里去)	3
总计		42

以上[+程度]构式只有“a+不+到+哪里+去”“能+a+到+哪里+去”可以出现在比较句中,其余构式都不能出现在比较句中。可以说,程度构式主要表“主观量”。此外,构式的意义往往是多层面的。例如,构式“a+不+到+哪里+去”除了表示“程度不是很高”外,有时还包含着一种反驳语气;构式“看+把+np+vp+得”“看+np+vp+得”有时有“过于vp”的含义,除了表示“高量”之外往往还包含着负面情感色彩。

值得指出的是,对于[+程度]义构式的形式特征乃至形成机制的探讨,有助于发掘更多的[+程度]构式,也有助于帮助计算机“举一反三”在碰到新的构式形式时,识别其是否具有“程度”义表达功能。现代汉语中表达程度最常见的手段是使用程度副词,但吕叔湘(1956:200-205)指出除了用程度副词之外,还可以用以下7种方式来表达程度^①:1)用数量词,例如“前天失迎,十二分抱歉。”2)感叹语气,例如“好快,一混就是十二年!”;3)叠用形容词,例如“好天分本来就**少**,有好天分又肯用功夫的真是**少而又少**”;4)用典型的事物来比拟,例如“一个白的白似雪,一个黑的黑似铁。”;5)用结果来衬托,例如“这场哭直哭得那**铁佛伤心,石人落泪**。”;6)用比较和假设来表示极限,例如“他的喉咙要**多高就多高**,他的中气要**多长就多长**。”;7)以含蓄表极致,例如“这一笔真是**说不出的妙**。”

吕先生的上述观察和分析对于认识构式“程度”义的形成机制(来源)有很大的参考意义。从其他表达范畴到表“程度”,涉及到多种不同类型的转喻来源:1)从数量范畴到表程度,如“一百个不放心”;2)从比较范畴到表程度,如“比专家还专家”;3)从否定范畴到表程度,如“别提多倒霉了”;4)从疑问范畴到表程度,如“你说倒霉不倒霉”;5)从感叹范畴到表程度,如“那叫一个帅啊”。当然,目前这样的概括还很难做到很周全,另外,针对计算机处理的需要,意义范畴的判别如何跟形式特征的分析相结合,也需要给出具体的规则。探讨程度构式的形成机制是一个大问题,需要另辟专文讨论,这里不再赘述。

5. <主体、对象、动因、后果、命令>要素的分析

上文给出了主观态度语义要素的7元组<主体,对象,类型,强度,动因,后果,命令>,其中类型、强度由构式自身决定,在上面第3、4节已经说明。而其余几个要素的抽取则需要分析构式所在的句子。下面对此展开讨论。

本文搜集的表主观态度义构式内部情况多样,因而并不是从任意一个构式所在的句子中都能抽取全部语义要素,针对一个具体的包含构式的句子,能抽取什么要素,主要取决于3个方面的因素:1)构式表达主观态度义的类型;2)构式的内部成分;3)构式的用法特点。

就构式表达主观态度义的不同类型来说,表“评价”义的构式一般只有<主体,对象>要素,句中不出现其他要素;表“情感”义的构式有<主体,对象,动因>要素,不出现<后果、命令>等要素;表“威胁”的构式有<主体,对象,动因,后果>要素,不出现<命令>要素;表“强令”的构式有<主体,对象,命令>要素,不出现<动因、后果>等要素。

就构式的内部成分来说,构式中变项的数量以及变项的范畴不同(如动词短语vp、名词短语np等)会影响要素的抽取,比如<动因>要素需要由np和vp共同来确定,如果构式中不含vp变项,就无法从构式中抽取到<动因>要素,这时就需要到构式所在句子的其他部分去寻找可能的要素成分。例如:

(19) 你看你,怎么这么晚才来!

例句中“你看你”是不含变项的构式,其中只包含<对象>要素“你”,这时候,需要将“你”跟后续小句中的vp成分“这么晚才来”合并构成<动因>要素。

就构式的使用特点来说,构式是否有“背责”(背后指责)用法,会明显影响<对象>要素的抽取。除[+评价]构式之外,[+主观态度]构式大多出现在对话语境中。如果构式只有“面责”(当面指责)的用法,那么<对象>就确定是“你”。比如威胁构式、强令构式只有面责用法,所以情感<对象>的取值都是“你”。

就一个特定构式而言,具体哪个成分对应哪种情感要素,需要从句法成分、构式内部成分的属性特征方面寻找限制条件。下面举例说明要素分析的一些规则。

规则1: [+评价]构式: 评价<对象>要素一般是句子主语,或者构式中的np变项。

大部分[+评价]构式充当谓语,这时句子主语是评价<对象>要素^②;少量[+评价]构式独立成句,这时构式中的np变项对应评价<对象>。例如:

(20) 老王动不动就讲创新。

(21) 亏他还是大学生。

在例(20)中,构式“动不动就vp”充当谓语,句子主语“老王”对应评价对象。在例(21)中,构式“亏np1还是np2”独立成句,构式的np1变项“他”对应评价对象。

规则2: [+情感]构式: 说话人为情感<主体>要素。构式中np变项如果是指人的,则为情感<对象>要素;构式中如包含vp变项,则跟指人np变项组合为情感<动因>要素。例如:

(22) 好你个李云龙!

(23) 我让你不听话!

(24) 老王他跑什么跑!

在例(22)中,构式的表人np变项“李云龙”对应情感<对象>,构式没有vp变项,无法从中抽取情感<动因>;在例(23)中,构式的np变项“你”为情感<对象>,vp变项“不听话”跟前面的np“你”组合,为情感<动因>。在例(24)中,构式的表人np变项“老王”对应情感<对象>,构式vp变项“跑”跟前面的np“老王”组合为情感<动因>,即说话人因为“老王跑”而对“老王”产生“不满”的情感。

规则3: [+情感]构式: 句子主语为情感<主体>要素。构式中的np变项为情感<对象>要素。句中无<动因、后果、命令>等要素。例如:

(25) 小王心里咯瞪一下。

(26) 小王拿他没办法。

在上两例中,句子主语“小王”对应情感<主体>。例(25)构式“心里咯瞪一下”没有变项,无法从中抽取情感<对象>、情感<动因>等要素;例(26)中,np变项“他”对应为情感<对象>。

规则4: 威胁构式: 前分句对应<动因>,后分句对应<后果>要素。例如:

(27) 任保老婆冲出来喝道“你个化石猴敢再讲,看我不要你的命!”(冯德英《苦菜花》)

例中构式的前分句“你敢再讲”对应<动因>要素,后分句“我要你的命”对应<后果>要素。

规则5: 强令构式: 构式中vp变项由自主动作动词(常见的如“过来、出去”等位移动词)充当,对应<命令>要素。例如:

(28) 你给我过来!

例中构式的vp变项“过来”对应<命令>要素。

值得一提的是,有些情况对于计算机的要素分析来说,难度比较大。比如:当构式变项是元语言(引语)用法,或者变项存在省略现象时,情感<动因>要素会难以分析为一个命题形式。例如:

(29) 你啊什么啊!

(30) 三块就三块!

在例(29)中,“啊”是元语言用法,即说话人对听话人说“啊”表示不满,情感<动因>是“你说‘啊’”。在例(30)中,“三块”是省略形式,完整的意义是“某物定价为三块”,即说话人对听话人“将某物定价为三块”表示不满,情感动因是“定价为三块”。计算机处理还难以得出这样的分析结果。

另外,上文列举规则时讨论的都是句子主语或构式变项与要素对应关系比较简单的情況。还有一些构式,存在两个以上成分组合对应一个要素的情况。例如:

(31) 武松放着公务员不当,偏要加入黑社会。

在例(31)中,情感<动因>为“武松+不+当+公务员”,由4个成分构成,涉及到句子主语,构式“放着+np+不+vp”中的变项np和vp以及常项“不”,同时还需要调整词序,才能得到这个完整的<动因>要素。这对计算机分析而言是相当困难的。

6. 结语

本文从情感分析的任务出发,讨论了跟情感表达有关的两类构式:[+程度]构式、[+主观态度]构式。其中[+程度]构式搜集到42条,[+主观态度]构式搜集到185条。从[+主观态度]构式包含语义特征的多少和类型来看,[+主观态度]构式可以只包含一种特征(例如只包含评价),也可以包含不止一种特征(例如包含情感和立场),有的[+主观态度]构式除了包含[+主观态度]特征之外还包含别的语义特征(例如果)。本文提出的情感分析相关特征拓宽了现有情感知识本体的覆盖面。

为了帮助计算机做文本情感分析,本文设立了主观态度语义要素7元组<主体,对象,类型,强度,动因,后果,命令>,并讨论了从句子中提取这些语义要素的规则。7个要素的设立和相应抽取规则的制定拓宽了文本情感分析要素抽取的范围。

为将本文研究切实地用于计算机文本情感分析任务中,下一步还将收集更多构式实例,构建情感义相关构式例句库,重点研究挖掘构式的句法语义特征以服务于[+主观态度]和[+程度]构式的识别及情感义要素的提取。

注释:

- ① 本文的例句,大部分自拟,小部分来自北大中国语言学研究中心CCL语料库(文中注明了具体出处)。
- ② 具体可以参看 http://www.keenage.com/html/c_bulletin_2007.htm
- ③ 具体可以参看 <http://ir.dlut.edu.cn/EmotionOntologyDownload> 和徐

琳宏等(2008:180-185)

- ④ 有关现代汉语构式知识库的详细情况,可访问 <http://ccl.pku.edu.cn/ccgd>。
- ⑤ 为方便针对计算机应用的形式描写,本文把构式分为“带变项构式”和“不带变项的构式”两种。前者如“要+多+a+有+多+a”,其中的a代表形容词性质的变项;后者如“心都提到嗓子眼了”,不包含变项。就“带变项构式”而言,如果变项词类单一,就用小写英文字母(如a,n,v等)表示;否则,就用大写字母X、Y、Z等表示。另外,为了醒目,在描述包含变项的构式时,一般在成分项之间用“+”起到区隔作用。
- ⑥ “接受”和“赞成”是有区别的,本文暂且把二者合并看作一个状态,为简略,一般用“接受”表示。
- ⑦ 引自侯玉波(2007:88)。
- ⑧ 例如图2输入句中的主语“他”对应情感要素中的“对象”;构式的常项是指构式中固定的成分,如图2例句中的“爱”“不”;变项是构式中可以替换的成分,如图2例句中的v“来”。这个句子抽取的“情感动因”要素值为“他不来”,是几个成分组合的结果。
- ⑨ 情感跟语言表达的主观性有密切的联系。语言主观性的研究主要集中在三个方面:说话人的视角(perspective)、说话人的情感(affect)、说话人的认识(epistemic modality)(沈家煊2001:269)。语言学有关情感的研究较多涉及“说话人的认识”即情态范畴,而文本情感分析领域的研究则侧重情感本身的分类。
- ⑩ 该构式的形式是“你+v+你+的+n”,跟下文例(17)构式形式相同,但是两者在形式以及语义上都存在明显差异。例(11)中的构式一般不省略“你”,而例(17)中的构式一般省略“你”;另外,两者在读音上也有明显差异,例(17)中的构式有非常明显的降调,而例(11)中的构式则没有。可参见Huang & Zhan(2015:630-635)。
- ⑪ 这7种方式是笔者根据吕叔湘(1956:200-205)9.87、9.88两小节总结而来。
- ⑫ 当构式包含两个或多个[主观态度]特征时,会出现不同特征的要素对应同一个成分的情况。例如,在例(25)中,“亏他还是大学生”同时包含[+评价]和[+情感]特征,“他”既是评价对象又是情感对象。为了称说简便起见,只将“他”说成评价对象。

参考文献

- [1] Goldberg, A. *Constructions: A Construction Grammar Approach to Argument Structure* [M]. Chicago and London: University of Chicago Press, 1995.
- [2] Huang, S. & W. Zhan. A construction approach to a personal marked construction in Mandarin Chinese [A]. In Lu, Q. & H. Gao. (eds.). *Chinese Lexical Semantics. CLSW2015 Lectures Notes in Artificial Intelligence, Vol. 9332* [C]. Cham: Springer, 2015: 625-637.
- [3] Thompson, G. & S. Hunston. Evaluation: An introduction [A]. In Hunston, S. & G. Thompson. (eds.). *Evaluation in Text: Authorial Stance and the Construction of Discourse* [C]. Oxford: Oxford University Press, 2000: 5.
- [4] 冯平. 评价论[M]. 北京: 东方出版社, 1995.
- [5] 侯敏, 滕永林, 陈毓麒. 评价短语的倾向性分析研究[J]. 中文信息学报, 2013(6): 103-109.
- [6] 侯玉波, 编著. 社会心理学(第二版) [M]. 北京: 北京大学出版社, 2007.
- [7] 刘鸿宇, 赵妍妍, 秦兵, 刘挺. 评价对象抽取及其倾向性分析[J]. 中文信息学报, 2010(1): 84-88, 122.
- [8] 吕叔湘. 中国文法要略[M]. 北京: 商务印书馆, 1956/2014.
- [9] 邵春. 英语评价型强势主位结构的双层评价功能[J]. 英语研

究, 2016(1): 67-75.

- [10]沈家焯. 语言的“主观性”和“主观化”[J]. 外语教学与研究, 2001(4): 268-275.
- [11]孙毅. 现代汉语情感动词事件致使—完成结构论[J]. 西北师大学报(社会科学版), 2018(4): 33-41.
- [12]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(2): 180-185.
- [13]夏雪, 詹卫东. “X什么”类否定定义构式探析[J]. 中文信息学报, 2015(5): 1-8.
- [14]詹卫东. 从短语到构式: 构式知识库建设的若干理论问题探析[J]. 中文信息学报, 2017(1): 230-238.
- [15]张国宪. 形容词的记量[J]. 世界汉语教学, 1996(4): 33-42.
- [16]赵妍妍, 秦兵, 刘挺. 文本情感分析[J]. 软件学报, 2010(8): 1834-1848.
- [17]郑娟曼. 现代汉语贬抑性习语构式研究[D]. 广州: 暨南大学, 2010.
- [18]中国社会科学院语言研究所词典编辑室编. 现代汉语词典(第5版)[Z]. 北京: 商务印书馆, 2005/2011.
- [19]周立柱, 贺宇凯, 王建勇. 情感分析研究综述[J]. 计算机应用, 2008(11): 2725-2728.

附录: 表主观态度义的构式及其语义要素分析示例

编号	构式基本信息		主观态度义要素				
	构式形式	构式实例	评价	情感	立场	后果	命令
1	是 + n	是地方	正面				
2	n1 + 是 + n1, n2 + 是 + n2	鼻子是鼻子, 眼睛是眼睛	正面				
3	够 + n	够朋友	正面				
4	不 + 怎么样		负面				
5	有点 + 那个		负面				
6	东 + X + 西 + Y	东看西看	负面				
7	七 + X + 八 + Y	七颠八倒	负面				
8	a1 + 一 + q, a2 + 一 + q	青一块, 紫一块	负面				
9	除了 + n + 还 + 是 + n	除了草还是草	负面				
10	np + 就 + 别 + 提 + 了	那事就别提了	负面				
11	不 + 是 + 我 + 说 + 你			不满			
12	你 + 看 + 你			不满			
13	你 + 说 + 你			不满			
14	真 + 是 + 的			不满			
15	让 + r + vp	让你不听话		不满			
16	好 + 你 + 个 + r	好你个李云龙		不满			
17	r + 啊 + r	老王啊老王		不满			
18	这 + 个 + r	这个老王		不满			
19	都 + 是 + np	都是老王		不满			
20	是 + r + 的 + 不是	是我的不是		不满			
21	r + 以为 + r + 是 + 谁	他以为他是谁		不满			
22	r + 算 + 老 + 几	他算老几		不满			
23	有 + r + 这样 + vp + 的 + 吗	有你这样办事的吗		不满			
24	r + v + r + 的 + n	你吃你的饭			赞同		
25	可 + 不 + 是 + 嘛			赞同			
26	谁 + 说 + 不 + 是 + 呢			赞同			

27	什么 + np	什么医生	负面	不满			
28	还 + np + 呢	还大学生呢	负面	不满			
29	亏 + np1 + 还 + 是 + np2	亏你还是一个大 大学生	负面	不满			
30	n1 + 不 + 是 + n1, n2 + 不 + 是 + n2	鼻子不是鼻子, 眼睛不是眼睛	负面	不满			
31	不 + n1 + 不 + n2	不男不女	负面	不满			
32	n1 + 不 + n1, n2 + 不 + n2	男不男, 女不女	负面	不满			
33	半 + X + 不 + Y	半阴不阳	负面	不满			
34	X + 就 + X	不来就不来		不满	赞同		
35	爱 + v + 不 + v	爱来不来		不满	不置可否		
36	爱 + vp + vp	爱怎么着怎么着		不满	不置可否		
37	随 + r + vp	随他怎么说		不满	不置可否		
38	v + 不 + v + 由 + 你	信不信由你		不满	不置可否		
39	vp + 什么	跑什么		不满	反对		
40	vp + 什么 + vp	跑什么跑		不满	反对		
41	哪 + 跟 + 哪 + 呀			不满	反对		
42	你 + 才 + np + 呢	你才半老徐娘呢		不满	反对		
43	我 + 把 + 话 + 放 + 在 + 这 + 里			不满	反对	+	
44	咱们 + 走 + 着 + 瞧			不满	反对	+	
45	看 + 我 + 不 + vp	看我不打死你		不满	反对	+	
46	再 + X + 我 + 就 + 不 + 客气 + 了	再说话我就不客气了		不满	反对	+	
47	一边 + 呆 + 着 + 去			不满			+
48	v + r + 的 + np	吃你的饭		不满			+
49	给 + 我 + vp	给我出去		不满			+
50	vp1 + 也 + 不 + 是, vp2 + 也 + 不 + 是	站也不是, 坐也不是			为难		
51	v + 也 + 得 + v, 不 + v + 也 + 得 + v	去也得去, 不去也得去			无奈		
52	心 + 提 + 到 + 嗓子眼儿 + 了				紧张		

说明: 后果、命令两列中的“+”表示对应构式有后果或命令特征。

基金项目: 本文系国家重点基础研究发展计划(973项目)“面向三元空间的互联网中文信息处理理论与方法”子课题4“融合三元空间的中文语言知识与世界知识获取和组织”(项目编号: 2014CB340504)和教育部人文社科基地2015年重大项目“构式的语义分析及其在语料库中的标注”(项目编号: 15JJD740002)的部分成果。

作者简介: 黄思思, 北京大学中文系博士研究生, 研究方向: 现代汉语语法、计算语言学。

詹卫东, 北京大学中文系教授, 博士生导师, 研究方向: 现代汉语语法、计算语言学。

责任编辑 孙毅