

# 文本情感分析中重叠现象研究

娜仁图雅  
首都师范大学  
nrty0910@cnu.edu.cn

徐晓音  
哈佛医学院  
xxu@bwh.harvard.edu

## 摘要

汉语中的重叠现象丰富，文本情感分析任务中应该密切关注语篇空间内的重叠现象及其交互状态。本文就重叠在文本中的样态、特点及情感标记功能进行了理论探讨；重点就构词性重叠、结构性重叠的表现形式及情感语义进行了分析；据此研究的基础上，本文就重叠现象在文本情感分析上的实际应用从几个方面进行了讨论。

**关键词：** 重叠现象；标记功能；构词形式；短语；应用

## A Study on Repetition in Text-based Sentiment Analysis

Tuya Naren  
Capital Normal University  
nrty0910@cnu.edu.cn

Xiaoyin Xu  
Harvard Medical School  
xxu@bwh.harvard.edu

## Abstract

Repetition of characters and words is a rich phenomenon in Chinese and worth close attention to the repetition usage and its interactions with other parts of text in the task of text-based sentiment analysis. This work discussed and analyzed the appearance, characteristics, and sentiment marking of repetition in text, with a focus on the analysis of the repetition of words and repetition of sentence structure, and the sentiment presentation given by the above repetitions. Based on the analysis, we elaborated on the practical application of identifying and using repetitions in text to assist sentiment analysis.

**Keywords:** Repetition, Sentiment marking, Phrase formation, Short phrases, Application

## 1 引言

近年来文本情感分析获得了很大的发展，并取得了广泛的应用。尤其是机器学习技术的发展，极大的提升了文本情感分析的准确率(赵妍妍2010)。比如Pang(2002)等人利用支持向量机在文本情感分析上达到了80%的准确率，Borbosa(2010)等人利用支持向量机分析推特上的文本情感，获得了81%准确率。Kamps(2004)等则利用WordNet的词向量来计算新词与代名词之间的语义距离来推断文本中形容词的情感倾向。其它方法包括有朴素贝叶斯算法(梁柯2019)和随机森林法(张月梅2020)(刘勇2019)。除了传统的机器学习方法，新技术尤其是深度学习算法也不断

被开发出来并应用在情感分析上,比如基于长短期记忆神经网络LSTM的情感分析(任勉2018)(卢强2019)以及在LSTM上发展起来的注意力机制的情感分析(王伟2019)(卢玲2018)。在上述现有的文本情感分析中,多数技术基于检测情感词来判断文字,语句,以及篇章文本中的情感极性和类别。周咏梅(2013)等构建了基于HowNet和Senti-WordNet的中文情感词典并考虑到情感词在不同语义环境下的情感倾向强度,同时使用词典校对方法对划分的情感词倾向强度进行了优化。比如Shen(2009)等专门建立了否定词、程度副词、感叹词和情感词词典来帮助计算机识别文本的情感倾向性。王素格(2009)等提出了一种先基于五种汉语情感词词表构建一个中文情感词表,然后利用该词表并采用线性加权组合的方法对句子情感进行分类的方法。黄高峰和周学广(2015)提出了一种基于词库权重、句式权重、修饰权重相结合的语句级细粒度情感倾向性分析算法,在这个算法中构建了一个细粒度情感词库,对其中的词语进行权重值的计算。张仰森(2018)等在对微博情感分析中则构建了一个包含情感词、程度副词、否定词、微博表情符号和常用网络用语的语料库帮助双重注意力模型对微博的情感分析。杨立公(2013)等提出了从文本颗粒度出发,总结了文本情感分析方法在情感词抽取,语料库和情感词典构建等几方面的发展。虽然情感词提供了一个明确的文字上的情感指示,但只依靠情感词来判断情感却有可能漏过其它表述情感的语言现象,进而影响情感分析的准确率。在这个文章里我们研究并讨论重叠作为一种语言现象在情感表达上的功能和如何通过检测重叠这种现象来增进情感分析的全面性和准确率。

重叠在情感意义的表达和交换过程中具有重要作用。能够使得情感信息在传递或被分析过程中,被更全面、更准确地解读出文本的情感意义。本文在计算语言的框架下,从面向文本情感分析的角度出发,对篇章文本中存在的各种重叠现象进行分析和再认识,实际上也是对情感表达各个组成成分或语义资源的分析与再整合。索绪尔(1980)提出“在语言状态中,一切都以关系为基础的”。重叠与情感表达之间所形成的关系,在文本情感分析任务中应注意以下两点:一是,重叠与情感表达之间的关系存在并协互动与互补关系。二是,重叠内部及外部的多个不同的重叠之间的情感语义关系,均具有共同的情感语义指向。在关系认识的基础上提取规则,能够帮助机器综合运用多种模态分析手段实现情感识别的目的。重叠现象在语篇中的表现形式多样,主要体现为说话人对某一事物或现象的内涵性评价或感受的深化,具有很强的情感表达和描述能力。重叠与情感的变化有直接或间接的关系。词语及小句两者重叠所处的句法位置能够带来情感表达(及强度)的差异,进而增加了文本情感分析的难度。当前的研究,尚欠缺对重叠现象的研究,也没有将重叠作为情感分析的一个判断指标,因此,值得对重叠现象进行深入细致的研究。

## 2 重叠样态及标记功能

文本情感分析各个层级的文本中,当出现重叠性的语言形式时,情感类型及强度都可能发生变化,具体表现为情感类型的凸显和情感强度的加深。原因在于重叠所形成的是共指表达,能够层层推进情感,所形成的情感语义潜势都指向情感类型的确定和强度的深化。

### 2.1 重叠样态

重叠是形成文本情感的诸多语言表现形式之一。文本中的词汇、句子、段落及篇章这四个层级中都有可能存在重叠性的语言信息。特别是词汇重叠极为常见,在不同层级的文本中都有分布。

#### 2.1.1 样态

文本中的语言表情形式是多样的,大体可分为内隐和显性的。汉语中的重叠现象,由于形式工整有规则可循,因此属于显性形式。重叠具有情感表达的功能,我们可以从其在文本中的普遍性和典型性样态来考察。例:

推开门一看,嗬!好大的雪啊!山川、树木、房屋,全部罩上了一层①厚厚的雪,万里江山变成了粉妆玉砌的世界。落光叶子的柳树上,挂满了②毛茸茸、③亮晶晶的银条儿;冬夏常青的松树和柏树,堆满了④蓬松松、⑤沉甸甸的雪球。一阵风吹来,树枝⑥轻轻地摇晃,银条儿和雪球儿⑦簌簌地落下来,玉屑似的雪末儿随风飘扬,映着清晨的阳光,显出⑧一道道五光十色的彩虹。《第一场雪·峻青》

文本情感分析中通常是以情感词作为主要评定指标,从该本文示例来看,含情感词“好、粉妆玉砌、五光十色”以及表情感的语气词“嗬、啊”。如果只是据此评定标准来计算该段文本的情

感的话,则显然不够充分,因为该文本中还含有8个重叠形式。这些重叠形式大体可归结为形容词的“AA式、ABB式”以及数量短语的“ABB式”。这些重叠是由几个意义相关或相近、相同或相似的结构,前后相连,经重叠,层层推进共同形成了一个与赞美、喜爱相关的情感表达链条。这些重叠形式在文本中的分布表明其实际上承担着情感表达的职能。该文本中的多个重叠与目标情感是个体与整体的关系。即文本的目标情感,是由包括重叠形式在内的诸多表情成分组织传送实现的。重叠产生的情感语义指向目标情感,当文本中出现多个重叠时,表示相互连接,共同推进和确立情感,表现为句或段落中的交集和交互状态。

### 2.1.2 特征

重叠形式的使用范围不限于词汇、句子、段落或篇章,其在文本中的分布可能不只局限于某个句子或某个段落当中,甚至可以跨越段际。其在文本中的使用样态来看,呈现出如下情感表达特征:(1)通常具有附加值。可使得情感表达的类型或强度在各种形式的重叠中发生变化,赋予更多的情感价值。(2)往往是临时形式。由于重叠形成了一个新的表达结构或单位,因此具备有限的能产性。(3)重叠的往往是凸显的。重叠的往往是可接受度高的语言成分,重叠后不影响对句子的理解,重叠的就是所要凸显的。(4)重叠形式有限制。重叠的形成受语言结构规则的限制,是建立在基本结构框架基础上的有规则的重叠。(5)具有共同的句法特征。重叠的目的是用以表示一个完整情感语义,句法上基本能够实现情感表达的自足。

## 2.2 标记功能

文本中重叠形式的使用具有情感标记功能。话语人在传递情感话语信息时,有意识地使用带有话语标记性质的重叠手段,实际上是对读者或听话人的语境假设加以限制和制约,其目的不在于改变话语所表达的语义内容或命题内容,而是要促进文本话语的理解和推进情感信息识别。一方面可以减少文本话语理解的阻力,另一方面则可推进情感信息的识别,提高读者或听话人理解文本的情感信息的效度,同时减少文本情感误读的出现,保证言语交际的正常进行。

### 2.2.1 情感强度加强标记

重叠所形成的情感强度潜势大于一个词或句子,是文本中情感强度被加强的指标之一。如果以零增为起点,定数每重叠或增加一次,便表示定数的增加,因此重叠具有程度加深的语义。例:

- a 我懂得母亲没有说完的话。妹妹也懂。我俩在一块儿,要①好好儿活……《秋天的怀念·史铁生》
- b 母亲扑过来抓住我的手,忍住哭声说:“咱娘儿俩在一块儿,①好好儿活,②好好儿活……”《秋天的怀念·史铁生》

上述两例来自同一作者、同一文章。a句中“好好儿”是“好”的“AA儿”式重叠。重叠后“好”具有了“殷切的要求或希望”的情感含义,形成了“好好儿活”这样一个全量的情感表达;b句中“好好儿活”出现两次,在全量的基础上实现定数的叠加,即情感语义及强度的再次叠加,因此情感强度明显高于a。进入到文本中,我们可以据此捕捉到从a到b的情感强度的变化。这是运算整合及语用推理的结果,所形成的特定的重叠性的结构,其语义特征是对原有结构特征的承继和发展(扩大),即大于原结构的情感语义特征,其潜在的情感强度大于一个简单句或复句,带来了一定的增值效应。因重叠而形成的结构化方式和特征,能够随着具体表达式的多次强化而调节情感表达强度。因此,重叠是文本中情感强度被加强的标记。

### 2.2.2 情感语义向心标记

语义指向是指句法结构中甲成分与乙成分有语义联系及语义所指的方向。文本中重叠的信息存在着一定的情感语义方向。重叠形式的使用其语义指向都应是向心的。如以下两种情况:(1)共指性重叠。是指句子中的重叠成分构成了共指表达,形成了联合指意(表情),能量增强,同时也意味着放弃其他指意可能。例:

看,①像牛毛,②像花针,③像细丝,密密地斜织着,人家屋顶上全笼着一层薄烟。《春·朱自清》“牛毛、花针、细丝”共指春天连绵的细雨。假设: X=雨; Y1=牛毛; Y2=花针; Y3=细丝; Z=细。无论是“牛毛”、“花针”还是“细丝”,都是为了说明春雨的“细”。“细”是“春雨”所具有的状态,即X、Y、Z是一种共指表达。将三个“比喻”辞格重叠

使用,使得三个喻体共享了本体的状态义素特征,同时排除其它特征,从而使得语义方向更加明确。

(2)非共指性重叠。是指句子中的重叠成分有分别由有交叉,结构上不是完全对应。非共指性的重叠同样能够将复杂的、强度高的情感展现出来。例:

它的颜色非常鲜艳。①头上的羽毛像橄榄色的头巾,绣满了翠绿色的花纹。②背上的羽毛像浅绿色的外衣。③腹部的羽毛像赤褐色的衬衫。《翠鸟·萧莽》

该例句中三个比喻小句的重复使用,不仅更具描述性,同时也附带有喜爱的情感。“头上的羽毛、背上的羽毛、腹部的羽毛”,是非共指性的本体,三个比喻小句的本体虽各有所指,却又共同指向“颜色非常艳丽”,因此非共指性的重叠仍然是具有语义向心力的。

综上,文本中的各个意义单元本身具有多种指意(表情)潜势。重叠的使用主要是出于所要表达的情感与强度形成匹配,意味着话语人力图实现表达的终极目的,具有优先实现指向核心情感的潜势,因此是语义向心标记。

### 2.2.3 情感动态发展标记

篇章文本中情感呈现点状分布,具有渐进性,且可跨过不同段落。经重叠形成的语义结构呈现出可操作性及整合性的特征,是一种动态性语义结构,能够充分展现人物(单个或多个人物)和事件的情感动态发展过程。例:

他们一队一队按照次序走,走过正对天安门的白石桥前,就举起灯笼火把,高声欢呼“毛主席万岁!”“毛主席万岁!”毛主席在城楼上主席台前边,向前探着身子,不断地向群众挥手,不断地高呼“人民万岁!”“同志们万岁!”《开国大典·李普》

该文本示例中,多个结构相同或近似的重叠(或反复)将欢腾的场面和不同人物的愉悦、热爱的情感不断推进,重叠的小句之间是递进关系。重叠作为一种情感信息组织的手段贯穿于文本之中,能够表达情感信息的动态发展及延续,是文本情感动态发展的标记。对其进行识别,可以据其自然段在文本进展中所处的位置为坐标(10%, 20%...90%, 100%)来分析每个自然段以及其中人物的情感的单独变化和各个人物相互的情感变化来分析总结文本中每个人物的情感基准、变化趋势和相对作用。

## 3 构形性重叠

汉语是语义型的语言,缺少形态变化,但是某些词类或个别词汇却存在构形上的变化,这种变化主要表现为重叠。“重叠之前的形式称为‘基式’,重叠之后的形式称为‘重叠式’”,我们称之为“构形性重叠”。构形性重叠形式的使用,往往能够用以表达喜爱、厌恶等情感类型,同时又与指小、增量形式指称联系在一起,以表示不同的情感类型和极性强度。

### 3.1 形容词重叠

有些性质形容词可以重叠,用以表示形状或程度的加深或适中,具有情感表达的功能,能够反映人的高强度的情感认识或体验。

#### 3.1.1 单音节重叠

单音节的重叠形式主要有“AA”、“AA的”、“AA儿”三种,如“弯弯、绿绿的、好好儿”。例:

- a 一位老人细细端详着毛主席,说:“这位首长,好像在哪儿见过。在哪儿呢?《毛主席在花山·翟志刚》
- b 小草偷偷地从土地里钻出来,嫩嫩的,绿绿的。《春·朱自清》
- c 我懂得母亲没有说完的话。妹妹也懂。我俩在一块儿,要①好好儿活……《秋天的怀念·史铁生》

上述例句经单音节重叠后,或多或少地都具有了情感表达的功能。a句里的“细细”将认真、仔细及疑惑的程度加强;b句则具有了喜爱的情感;c句体现为一种殷切的希望。另外,两个单音节能够联合重叠,形成AABB式,这两个单音节词汇之间互为反义或近义,通常用来表达话语人的对某人某事的概括性的评价和认知。例:

- a 这时，又从上边的河湾湾里漂来了好多大大小小的蒲团儿，...。《小翠鸟打房子·王季》
- b 左面山坡上高高低低一幢一幢土红色和灰色的三层楼房吸引着我的眼光。《MV旧的春天·巴金》

例句中的重叠是单音节的“大、小”，“高、低”联合形成的重叠，表达的是话语人对事物的概括性一种认识，具有指多增量的情感语义。

### 3.1.2 双音节重叠

双音节重叠主要有“ABB”、“AABB”、“BAA”三种，如“红彤彤、曲曲折折、红通通”。例：

- a 没有月光的晚上，这路上阴森森的，有些怕人。《荷塘月色·朱自清》
- b 店主人高高兴兴地捧出了笔砚纸张。《文天祥·沈起炜》
- c 警卫员只好接过茶叶筒，端端正正地向毛主席敬了个礼。《毛主席在花山·翟志刚》

a句式用了ABB式重叠，渲染了周遭阴森之感，“阴森”本身就是负面情感词，重叠后在此处表达了一种更加强烈的恐惧；b句使用的是AABB式重叠，“高兴”是正面情感词，本身强度就高，重叠后其所表达的仍然是高兴的情感，但强度却达到最高；c句中使用的是“端正”的AABB式重叠，表达了警卫员对毛主席的敬重的情感。所以说，AABB式重叠可以表达一时一地的态度，这也是双音节重叠的一个情感语义特点。

### 3.2 名词重叠

名词重叠式主要有“AA”，“ABAB”，“AABB”和“ABB”式。重叠之后名词一般具有赞同、喜爱、亲昵等(与认同类相关的)情感语义。例：

- a 人人都有饭吃，家家都有房住。
- b 这孩子猴精猴精的。
- c 我爱家乡的山山水水，我就想用我自己学到的知识去把它变得更美好。《你怎么也想不到·路遥》
- d 可是从此以后，每逢看见蜜蜂，感情上①疙疙瘩瘩的，总不怎么舒服。《荔枝蜜·杨朔》
- e ①青线线(那个)②蓝线线，蓝格英英(的)彩，生下一个③兰花花，实实的爱死人。《兰花花·民歌》

a句的增加了“所有、每一”的含义，是对“每个人、所有家庭”的评价，一种增量指称；b句的“猴精”是名词，重叠后语义发生变化，用以表示“机灵、聪敏”的含义，是指小形式指称。表达的是与“夸赞”相关的情感类型，且情感强度很高；c句“山山水水”是单音节名词联合重叠所形成的AABB式，具有繁多的含义，凸显了对家乡热爱的情感；d句“疙瘩”一词原本没有情感，经过重叠而具有了情感表述的功能，用以表达“不快”这一情感类型，是一种增量形式指称；e句是陕北方言样本，含有多个名词的重叠，使得兰花花这一人物形象跃然纸上，表达的是“喜爱”的情感，属于增量形式指称。

### 3.3 动词重叠

汉语中有些动词本身就含有情感义，如“喜欢、祈盼、虐打”等；而有些则不含，如“看、听、说”等。同样的，有些动词能重叠，有些则不能。如某些用来表示动作行为的动词就能够发生重叠，并有了表短时或轻微的含义。其重叠形式，单音节动词为AA式，如“想想”；双音节动词为AABB式，如：“研究研究”。例：

- a 织云抬起眼光，看看远远的蓝天。《我们的歌·赵淑侠》
- b 贾芹走进书房，只见那些下人指指点点，不知说什么。《红楼梦·第93回》

- c 她也笑了，坐在我身边，絮絮叨叨地说着：“看完菊花，咱们就去‘仿膳’，你小时候最爱吃那儿的豌豆黄儿。《秋天的怀念·史铁生》

重叠后的动词会被赋予情感语义。a句中的“看”本身不具有情感表达功能，经AA式重叠具有了短时的含义；b句中的“指点”经重叠后发生了语义上的改变，具有了贬义；c句“絮叨”一词具有负向的情感，重叠形式“絮絮叨叨”不仅具有指多的含义，同时加深了反感的情感强度。此外，有些动词的AA式重叠使用场景比较固定，如“红旗飘飘、蝴蝶飞飞”等，表示的是动作的反复和延长，而不是短时或轻微的含义。

### 3.4 数词重叠

汉语中可进行重叠的数词非常有限，最常见的数词重叠形式是“AA”式和“AABB”式。其中“一、九、千”可重叠为“AA式”；“二、三、千、万”可重叠为“AABB式”。例：

- a 心似双丝网，中有千千结。
- b 九九八十一难。
- c 天上五颜六色的火花结成彩，地上千千万万的灯火一片红。《开国大典·翟志刚》
- d 问今是何世，乃不知有汉，无论魏晋。此人一一为具言，所闻皆叹惋。《桃花源记·陶渊明》
- e 等到得不着票子，便不免有了三三两两的怨声了。《旅行杂记·朱自清》

a、b句中的数词经重叠后具有指多增量的含义，表示心结多、磨难多。句子情感强度被增强；c句则具有了指数增量的含义；d句数词“一”重叠之后意义和功能均发生改变，不仅用以指多，还具有逐一地，详细、详尽地对某人某事或某物的进行介绍的含义，多用于动词前，具有正向的情感，情感强度被增强；e句中“三、两”本身就是指少数词，重叠后具有了零散的语义。需要说明的是：AABB式中只有这一例是减量指少，情感强度被削弱。其它的“AA式”或“AABB式”都是指多增量，情感强度被增强。数词重叠带来的词义增加。重叠后的数词除了“三三两两”具有指少减量的语义之外，其他赋予“指多增量”的语义。数词重叠可使得句子的情感强度增强或削弱，因此具有情感表达功能。

### 3.5 量词重叠

量词重叠不仅增加了意义，还增加了功能。李宇明(1996)认为AA式是量词重叠形式。单音节量词大多可以重叠，如“个个、件件、条条”等。量词重叠之后增加了“每一、逐一”或“多”的含义。例：

- a 个个都是好样的。
- b 我家洗砚池头树，朵朵花开淡墨痕。《墨梅·王冕》
- c 层层叶子中间，零星地点缀着些白花，有袅娜地开着的，有羞涩地打着朵儿的。《荷塘月色·朱自清》

上述例句的重叠之后，增加了“每”和“多”的含义。a句表达了对每一个人的赞赏；b句则增加了繁多的语义，表达的是诗人对梅花的喜爱之情；c句则增建了逐一的含义，这些量词的重叠使得被描述的对象更加形象化和可观可感。

综上所述，主要是就形容词、名词、动词、数词、量词的所形成的重叠进行了分析。重叠形式的使用在情感表现力上，增量形式指称的特点较为突显，具有程度更深、更细腻的特点；在语言使用体验上，生动而形象，能够呈现出更多的情感表达效果，且语用接受程度高。

## 4 结构性重叠

文本情感分析不仅要关注语言系统中稳定的、规则的、典型的语言现象，还应关注特殊的、偶发的、创新性的塑造语言结构的现象。结构性的重叠，可视为特殊的、创新性的情感表达手段，主要表现在比况短语、“的”字词组重叠、数量短语、状中短语及定中短语的重叠上。

#### 4.1 比况短语重叠

比况，实际上是比喻的灵活多样用法之一。本体可不出现，喻词及喻体出现，重叠排列成结构相似或相同、互相映衬的平行结构(或句式)来表达话语人的思想情感。比况结构的重叠使用，将使得句子形象生动、情感类型更加明确和凸显，情感强度高。例：

- a 你①像雾②像雨又③像风。
- b 野花遍地是：杂样儿，有名字的，没名字的，散在草丛里，①像眼睛，②像星星，还眨呀眨的。《春·朱自清》
- c 有的①像峰峦，②像河流，③像雄狮，④像奔马……它们有时把天空点缀得很美丽，有时又把天空笼罩得很阴森。《看云识天气·朱泳》

a句是三个比况联合形成的比况结构重叠，将人比作“雾、雨、风”，表达的是话语人为之着迷的情感；b句有两个比况结构，把野花人格化，使它们具有人的情感，形成的拟人性的表述，用以表达作者强烈的喜爱的情感。c句的喻体由三个比况结构重叠构成，将云比作是“峰峦、河流、雄狮、奔马”，表达的是作者对云的形态变幻的认识，同样是一种与喜爱相关的情感。

#### 4.2 “的”字短语重叠

“的”字短语由助词“的”附着在实词或短语后组成，用来指称人或事物，属于名词性短语，其功能相当于一个名词，如“大的、卖菜的、跑堂的”。其重叠使用通常是话语人为了进行详尽地介绍而采用的一种表达方式，是增量指称，对小句情感强度具有补强作用。例：

- a 夜市上好不热闹，有①摆摊的、②卖药的和③赚吆喝的……
- b 我用手拨开草一看，原来青草下边藏着满满一层小花，①白的、②黄的、③紫的。《花的勇气·冯骥才》
- c 我们继续拍掌，树上就变得热闹了，到处都是鸟声，到处都是鸟影。①大的、②水的、③花的、④黑的，有的站在树枝上叫，有的飞起来，有的在扑翅膀。《鸟的天堂·巴金》

上述例句大都是三个以上“的”字短语的连续使用，我们视此种语言表达形式为“的”字短语重叠。“的”字短语本身不具情感意义，经重叠后其作用主要是罗列，增加了指多的含义，因此具有间接性地补强句子情感的作用。其中，a句“的”字短语的重叠主要是对前文夜市的热闹程度的补足和加强；b句重叠的目的不单是罗列，同时也是希望给读者以身临其境之感，间接表达了作者对“满满一层小花”的喜爱；c句通过“的”字短语的重叠将鸟的种类颜色等进行了大致的介绍，目的是以此保持与“鸟的天堂”这一主题及整体情感语义上的相和谐，表达的是对奇特幽美和别具洞天的“鸟的天堂”的赞美。

#### 4.3 数量短语重叠

通常量词与数词组成数量词组产生重叠，形成：一AA式、一A一A式，如“一件件、一本一本”。具有“繁多”和“按某种方式或状态进行”的含义。两种形式可互换使用。例：

- a 月亮越升越高，穿过(一缕一缕)轻纱似的微云。《月光曲·杨爽》
- b 玉屑似的雪末儿随风飘扬，映着清晨的阳光，显出(一道道)五光十色的彩虹。《第一场雪·峻青》
- c 白云来了，(一大团一大团的，)从祖父的头上飘过，好像要压到了祖父的草帽上。《祖父的园子·萧红》

a、b句中的数量词组的重叠，具有指多增量的功能，这种指多表现出的是与喜爱或认同等相关的情感。在情感表达上，如果省略例句中括号中的重叠，比如a与原句相比较，我们可以体会到“一缕一缕”的使用更加强了作者轻松平静的心态。同样的，把b句中的重叠去掉便会降低作者对雪欣赏赞许的强度。c句中的“一大团一大团”则凸显了云的状态，增强了作者对白云这样一个具正面极性事物的倾向性看法，如果将其去掉，便觉少了轻松愉快的情绪，虽还保有正面的极性，却更偏近于中性了。

#### 4.4 状中短语重叠

状中短语由状语和中心语组成。该结构的重叠形式丰富多样，能够表达认同、赞美、喜爱等情感，经重叠后的状中结构能够赋予句子更高的情感强度。例：

- a 小鱼儿在荷叶下笑嘻嘻地游来游去，捧起①一朵朵②很美很美的水花。《荷叶圆圆·胡木仁》
- b 昙花开放的声音是短促的，茶花开放的声音是悠长的，不管短促或悠长，都是①那么动听，②那么迷人。《花开的声音·爱玲》
- c 他们正用劳力建设自己的生活，实际也是在酿蜜——①为自己，②为别人，③也为后世子孙酿造生活的蜜。《荔枝蜜·杨朔》

a句中含有两项重叠，其中②是程度副词和形容词构成的状中结构的重叠，联合加强了“美”的程度和状态，表达的是喜爱的情感；b句是代词和形容词构成的状中结构的重叠，表达的是赞美的情感；c句是由介词短语和形容词构成的状中结构的重叠。该句当中不含情感词，假设只有一项状中结构的条件下，该句是不含情感的。该句中含有三项状中结构，其重叠使用赋予了句子颂扬的情感。

#### 4.5 定中短语重叠

定中短语由定语和中心语构成，两者是修饰关系，修饰语一般由形容词和名词充当，中心语是被修饰的对象。修饰语和中心语之间有时用“的”来连接。定中短语是不自由短语，通常不能独立使用，只能做主语宾语。其在文本中有连续反复使用的情况，主要体现为对主语或宾语的凸显，凡是重叠的就是凸显的，因此，能够赋予句子情感。例：

- a 小河是娴静的，宛如明镜一般，倒映着①红色的花、②绿色的树。《家乡的小河·佚名》
- b ①青的草，②绿的叶，各色鲜艳的花，都像赶集似的聚拢来，形成了光彩夺目的春天。《燕子·郑振铎》
- c ①秋的味，②秋的色，③秋的意境与姿态，总看不饱，尝不透，赏玩不到十足。《故都的秋·郁达夫》

定中短语在文本中的重叠使用，意味着修饰性词语出现的几率越高，修饰性的词语通常具有较强的形象色彩及情感色彩，因此实现情感表达的可能性就越大。a、c两句重叠的使用，不仅限定了对象，同时具有很强的画面感。因此，a句凸显了小河倒影的美感，b句的重叠则与“光彩夺目的春天”相匹配，表达出对春的赞美；c句由三个定中结构反复使用，联合形成一个大主语，修饰语及中心语之间具有领属关系，因此是对主体对象的凸显。相比之下，用“秋天”来替换做主语也是可以的，但却丧失了主语所承担的情感功能。因此定中短语重叠具有赋予句子情感作用。

综上所述，句子中的结构性重叠现象能够用以表达情感。董秀芳(2016)认为“词组的重叠是一种句法上的重叠，实际上是同义并列。这种手段所造成的形式都有描摹状态的意味，这种描摹带上了加强程度的含义，平行于词法重叠中的增量功能，而这种增量的描摹功能也是取决于说话人的认识，因此也具有明显的主观性。”主观性是实现情感表达的必要因素，因此，在文本情感分析中结构性重叠具有情感表达的功能。

### 5 重叠在情感分析上的应用

在前面几节里我们总结了重叠的几种表现形式。可以看出来，在情感表达上，重叠不仅凸显了情感阶段的出现而且加强了情感表达的强度和细腻化，另外，重叠也把一些中性词变的有了情感色彩。例如，“冷”本身不具有感情色彩，重叠后“冷冷地”则表现了负面情感。类似的，“热”作为单独一个字是中性的，当它与虚词“乎”结合成“热乎乎”就变成了一个ABB形式的正向表达，同理，“热热闹闹”就变成了一个AABB式的正面情感表达。因此，通过对重叠现象的分析，我们可以提取一些规律来帮助计算机识别情感，从而可以从几个方面来有效地提高文本情感分析的准确性。



第一个方面就是通过检测字词尺度上的重叠。根据上述的讨论, 我们看到重叠的有效长度有不同的尺度, 既可能以重复单个字或单个词为单元的形式出现, 如AA或ABAB的形式, 亦或AABB和ABCABC形式; 也可能以重复短语或短句为单元的形式出现。在以字词为尺度的重叠中, 我们可以设置算法来寻找AA或ABAB形式的语言现象, 如采用一个带记忆的滑动窗口(moving window) $w$ 来检测AA或ABAB。在这个过程中, 假设 $w$ 的记忆长度为 $k$ , 在位置 $n$ 的输入为 $w(n)$ , 则算法检测是否

$$w(n) = w(n-1) \quad (1)$$

来检测AA式的重叠; 同样的, 算法检测是否

$$w(n) = w(n-1) \quad (2)$$

$$w(n-2) = w(n-3) \quad (3)$$

并

$$w(n) \neq w(n-2) \quad (4)$$

来判断AABB式的重叠; 以及检测是否

$$w(n) = w(n-2) \quad (5)$$

$$w(n-1) = w(n-3) \quad (6)$$

来判断ABAB式的重叠。同时我们还需要考察重叠的组合用法, 如在例句“毛茸茸、亮晶晶的银条儿”出现的ABB, CDD形式的重叠, 那么这个时候就需要把ABB, CDD作为一个整体考虑来判断其表述的情感极性和强度。

第二个方面则检测短语或短句尺度上的重叠。实现短语或短句的重叠检测需要设计灵活的算法在检出句式结构的同一性的基础上准确地检测语义上的重复或类似性, 比如例句“漓江的水真静啊, 静得让你感觉不到它在流动; 漓江的水真清啊, 清得可以看见江底的沙石; 漓江的水真绿啊, 绿得仿佛那是一块无瑕的翡翠。”算法可以设计成在一个长度为 $K$ 的记忆片段中, 以标点符号为隔断定义短句, 在当前的第 $n$ 个短句的位置上首先分析短句的结构, 如动宾结构, 主谓结构, 或动状结构, 然后搜索第 $(n-1), (n-2), \dots, (n-j)$ 个短句知道第 $(n-j-1)$ 个短句的结构与当前短句的结构不同式停止。这时我们可以判断从第 $(n-j)$ 到第 $n$ 个短句具有相同的结构。下一步我们可以设计算法判断每一个短句的情感极性和类别。

第三个方面则需要针对不同尺度的重叠进行综合整理。在检测分析字词尺度和短句尺度上的重叠现象的同时, 算法还需要能够处理复杂的重叠现象, 比如以下例句中出现的嵌套式的重叠, “一会儿红彤彤的, 一会儿金灿灿的, 一会儿半紫半黄, 一会儿半灰半百合色”。在这个例句里我们既观察到短句结构上的重叠, “一会儿.....”, 同时还可以看到在第一和第二个短句出现ABB式的字词重叠, 即“红彤彤”“金灿灿”, 而在第三和第四个短句中出现混合ABAB式的字词重叠, 即“半紫半黄”“半灰半百合色”。在这种语言现象中, 我们需要设计算法使其正确解读“红彤彤”“金灿灿”隶属于短句尺度上的重叠, 用来表示同一个情感场景, 而不是两个分别的情感。如果我们用一个矢量来代表短句尺度的情感表达, 针对上述例句我们可以用四个矢量分别代表其中的短句, 如 $F_1$  = “一会儿红彤彤的”,  $F_2$  = “一会儿金灿灿的”,  $F_3$  = “一会儿半紫半黄”,  $F_4$  = “一会儿半灰半百合色”。而其中 $F_1$ 又包含一个字词尺度的情感表达 $f_1$  = “红彤彤”,  $F_2$ 包含 $f_2$  = “金灿灿”。同样的,  $F_3$ 包含了 $f_3$  = “半紫半黄”,  $F_4$ 包含有 $f_4$  = “半灰半百合色”。在情感识别上,  $F_1$ 到 $F_4$ 的短句重叠可以指示这是一个富有情感的表述, 但仅凭短句重叠还不能判断情感极性, 当结合了 $f_1$ 到 $f_4$ 的字词重叠, 特别 $f_1$ 和 $f_2$ , 我们就可以判断出这是一个正面的情感极性。在这个例子里我们还可以观察到重叠现象之间存在的互相修饰作用, 亦即虽然 $F_3, F_4$ 中的“一会儿半紫半黄, 一会儿半灰半百合色”本身是中性的表述, 但结合了 $F_1$ 和 $F_2$ 的正面情感极性后, 我们就可以判断 $F_3$ 和 $F_4$ 也变成了或参加了正面情感的表述。而这种通过短句重叠来赋予中性表述以极性的功能在依据情感词来表述情感的文句是罕见的。

第四个方面在于准确判断重叠在单独使用时赋予文本的情感含义。根据前面的分析和举例, 我们注意到在没有情感词的情况下, 重叠可以单独表达情感, 如“实际也是在酿蜜——为自己, 为别人, 也为后世子孙酿造生活的蜜”中, 重叠的状中结构将原本是中性的“实际也是在酿蜜”变成了正面的赞美。这种情形如果只是检测词库里的情感词很可能无法判定句子的情感属性, 而

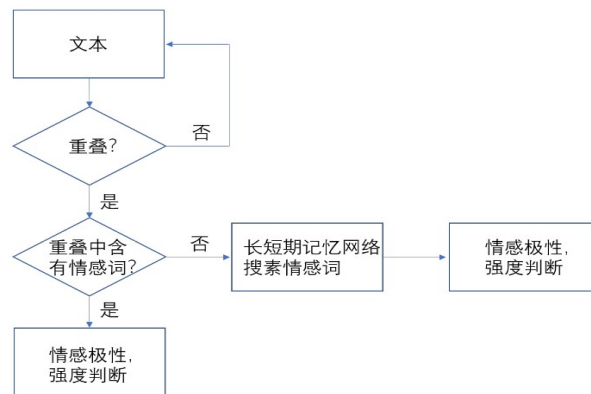


Figure 1: 重叠现象的检测与情感判断

当我们设计一个算法检测出这一重叠时，算法便可以向前和向后搜索到“酿蜜”和“生活的蜜”来综合考虑一个重复强调“酿蜜”的句子应该是含有正面情感的。从这个方面来考虑，现有的情感分析方法包括深度学习方法可以在情感词的基础上拓展对情感表现的定义，将重叠作为一种表达情感的现象包括在训练和验证过程中。从计算语言学的角度出发，根据重叠在情感表达上的多样性，我们可以设计类似于图1的计算流程来首先检测重叠是否出现，当检测到重叠时，须进一步判断重叠的语形本身是否为情感词，以及重叠语形长度附近是否与情感词存在共现的状况。如果含有情感词，即视为共现，可以直接据情感词的标注状况来判断它所表现的情感极性和强度。如果不含，即非共现，则可通过机器学习结合情感词来判断整个重叠现象的情感极性和强度。在这一步需要记录与重叠相关的情感词的位置，因为情感词可能出现在重叠之前或之后，所以可以利用长短期记忆网络或类似的具有记忆功能的机器学习技术来捕捉到情感词和重叠直接的关联并加以分析。例如，在长短期记忆网络中，根据重叠的表现形式我们可以用来加强不同长度记忆的强度。综合以上几个方面的讨论，我们可以得出结论，重叠作为一种语言现象，有着使用上的多样性和增进情感表达的作用，既可以与情感词联合使用也可以单独使用。对重叠的准确检测和正确分析可以有效地提高自动化的情感分析的表现。在深度学习被广泛用于文本分析的情景下，文本中检测出的重叠现象可以用于加强对神经网络的训练。

## 6 总结

本文中我们从重叠的语言现象出发，探讨和归纳整理了不同形式和尺度上的重叠现象，既有最简单的单音节字的重叠，也有词组的重叠，还有复杂的短语和短句的重叠。通过分析，我们发现重叠对文本的情感表达有着多方面的作用，重叠既可以与情感词共现配合来显性的表达情感，也可以单独出现来隐性的表达情感。在这种情况下，通常需要结合上下文才能准确判断重叠所表现的情感。从文本情感分析视角来看，重叠现象可以作为一个情感表达的标记，可以据此设计自动分析方法检测文本中的重叠，当检测到时可首先预期正面或负面的情感表达，通过对重叠的细致分析，再进一步细致地判断情感的极性和强度。重叠对于文本情感分析的影响，一是，既能突出情感表达的语句，也能起到调控情感强度的作用，同时，重叠还能弥补仅仅依靠情感词进行情感分析的一些不足。二是，重叠，尤其是句式结构上的重叠，以及当重叠不依赖情感词而表达情感时，对情感分析也是一个新的挑战。如何把重叠现象和情感词相结合以开发出更好的情感分析技术还是一个新方向，本文对此做了一些探讨。在未来的研究中，我们计划对重叠现象做进一步的分析，归纳总结如何对重叠进行自动化的计算分析，并与情感词语料库结合起来设计具有针对性的算法，并于实验相结合来验证算法的表现。

## 致谢

本文中娜仁图雅的工作为北京市人力资源与社会保障局外派项目“口述文本的关键语义信息提取(2018PC03)”的阶段研究成果；同时得到国家社科基金重大委托项目“语言大数据挖掘与文化价值发现”(14&ZH0036)的资助。

## 参考文献

- 赵妍妍, 秦兵和刘挺. 文本情感分析. 软件学报, 21(8):1834-48.
- Pang B, Lee L, and Vaithyanathan S. Thumbs up? Sentiment classification using machine learning techniques. *Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing*, 10:79-86.
- Borbosa L and Feng J. 2010. Robust sentiment detection on Twitter from biased and noisy data. *Proceedings of the 23rd International Conference on Computational Linguistics, Association for Computational Linguistics*, 36-44.
- Kamps J, Marx M, Mokken RJ, and De Rijke M. 2004. Using WordNet to measure semantic orientations of adjectives. *In LREC*, 4:1115-1118.
- 张月梅和刘媛华. 2020 基于K近邻和随机森林的情感分类研究. 计算机与数字工程, 48(2):367-371.
- 刘勇和兴艳云. 2019 基于改进随机森林算法的文本分类研究与应用. 计算机系统应用, 28(5):220-225.
- 梁柯, 李健, 陈颖雪和刘志钢. 基于朴素贝叶斯的文本情感分类及实现. 智能计算机与应用, 5(34).
- 任勉和甘刚. 2018. 基于双向LSTM模型的文本情感分类. 计算机工程与设计, 39(7):2064-2068.
- 卢强, 朱振方, 徐富永和国强. 2019 融合语法规则的Bi-LSTM中文情感分类方法研究. 数据分析与知识发现, 3(11):99-107.
- 王伟, 孙玉霞, 齐庆杰和孟祥福. 2019. 基于BiGRU-Attention神经网络的文本情感分类模型. 计算机应用研究, 12:8.
- 卢玲, 杨武, 王远伦, 雷子鉴和李莹. 2018. 结合注意力机制的长文本分类方法. 计算机应用, 38(5):1272-1277.
- 周咏梅, 杨佳能和阳爱民. 2013 面向文本情感分析的中文情感词典构建方法. 山东大学学报(工学版), 43(6):27-33.
- Shen Y, Li S, Zheng L, Ren X, and Cheng X. 2009. Emotion mining research on micro-blog. *In 2009 1st IEEE Symposium on Web Society*, 71-75.
- 王素格, 杨安娜和李德玉. 2009. 基于汉语情感词表的句子情感倾向分类研究. 计算机工程与应用, 45(24):153-161.
- 黄高峰和周学广. 2015. 一种语句级细粒度情感倾向性分析算法研究. 计算机应用与软件, 32(4):239-42.
- 张仰森, 郑佳, 黄改娟和蒋玉茹. 基于双重注意力模型的微博情感分析方法. 清华大学学报(自然科学版), 58(2):120-130.
- 杨立公, 朱俭和汤世平. 2013 文本情感分析综述. 计算机应用, 33(06):1574-607.
- 索绪尔. 1980. 普通语言学教程. 北京:商务印书馆.
- 李宇明. 1996. 论词语重叠的意义. 世界汉语教学, 1(11):10-19.
- 董秀芳. 2016. 主观性表达在汉语中的凸显性及其表现特征. 语言科学, 15(6):561-70.