

文章编号: 1000-5862(2013)02-0130-06

# 情感图式构造及其在文本情感计算中的应用

任巨伟, 杨 亮, 林鸿飞\*

(大连理工大学信息检索研究室 辽宁 大连 116024)

**摘要:** 在认知评价理论的基础上, 将情境刺激及评价行为映射到多级概念空间中. 借助 HowNet 和同义词林, 使用基于模式的上下位关系获取方法, 构造了 4 种情感图式. 结合认知语境和语用推理, 将其运用于句子级文本情感计算中. 实验结果表明: 情感图式可以有效地辅助文本情感分析, 提高分析准确率.

**关键词:** 情感图式; 情感计算; 概念空间; 认知语境; 语用推理

**中图分类号:** TP 391

**文献标志码:** A

## 0 引言

情感计算, 按照美国 MIT 实验室 Picard 教授的著作《Affective Computing》中的描述, 被定义为与情感相关, 来源于情感或者能对情感施加影响的计算<sup>[1]</sup>. 所涉及的领域包括计算机科学、心理学和认知科学<sup>[2]</sup>. 随着网络信息的爆炸式增长, 文本情感计算研究逐渐成为研究热点, 但目前情感计算方法局限于传统的文本分析技术. 从认知学和情绪心理学角度的研究方法才刚刚兴起, 认知科学和情绪心理学体系也可以作为潜在有益的研究思路.

目前, 文本情感计算主要使用文本分类、情感词加权和语义模式等文本情感分析技术. 句子级文本情感计算的典型方法是文本分类方法或者结合情感本体和语义特征的方法<sup>[3]</sup>, 后者从句子的词汇和结构两方面考虑, 提取影响语句情感色彩的语义特征, 最终确定句子在其所处文本大范围语境中的情感色彩, 但是该方法对于没有显性情感词标识和语义特征的文本无法判断其情感.

考虑到情绪和认知的密切关系, 徐琳宏等<sup>[4]</sup>从认知学和情绪心理学角度, 提出一种文本情感认知模型, 借助情感图式提高文本情感识别的准确率. 这种情感图式的构造主要以手工标注和外部词典扩展为主, 没有自动抽取步骤, 且其规模和灵活性有待提

高; 另其实现文本情感分析的方式主要以关键字匹配和词语搭配为主, 缺少一定的推理功能.

本文在徐琳宏等提出的情感图式的基础上提出了一种更加规范化的情感图式构造方法, 使用基于模式的上下位关系获取方法, 结合认知语境和推理理论, 扩大了图式规模, 增加了图式的推理功能, 并将其应用于句子级文本情感计算中, 以辅助传统的基于情感词和语义特征的方法进行情感分析.

## 1 理论基础

### 1.1 情感和认知

情绪是多成分组成、多维量结构、多水平整合, 并为有机体生存适应和人际交往而同认知交互作用的心理活动过程和心理动机力量<sup>[5]</sup>. 情感多是指社会性高级感情, 是具有稳定而深刻的社会含义的感情性反应. 为表述方便, 本文中“情绪”和“情感”不作区分, 统称为“情感”. 采用国内外通用的 Paul Ekman 情感分类方法<sup>[6]</sup>, 将情感分为 6 大类: “喜”, “怒”, “哀”, “惧”, “恶”, “惊”. 认知或认识在心理学中是指通过形成概念、知觉、判断或想象等心理活动来获取知识的过程, 即个体思维进行信息处理的心理功能.

阿诺德指出, 情绪的发生需要有 3 个要素, 即认知评价、生理过程和环境. 情绪产生在于情境刺激,

收稿日期: 2012-11-15

基金项目: 国家自然科学基金(61272370, 60973068)和辽宁省自然科学基金(201202031)资助项目.

通信作者: 林鸿飞(1962-), 男, 辽宁大连人, 教授, 博士生导师, 主要从事搜索引擎、文本挖掘、情感计算和自然语言理解等方面的研究.

而当前的情境刺激有什么意义和作用,都需要通过认知评价来揭示.拉扎勒斯认为,“情境刺激—评价—应付”是一系列连续的社会行为,对同一情境刺激采取不同的评价和应付方式会唤醒不同的情绪.情绪是认知评价的结果,认知评价理论已成为情感图式构造的理论基础.

## 1.2 情感图式

认知研究表明,在人对客观事物的认知活动中,认知过程的思维反映在不同的思维反映层次中,而图式的层次化模型能较好地表征这一思维反映过程,即表征人类对于新旧事物的理解和认知模式.图式概念最初在1932年由英国心理学家 Frederic Bartlett 引入到心理学和教育学领域.它是一种知识的抽象结构,该结构隐含了不同组成概念间的关系<sup>[7]</sup>.图式可表示为一系列具有语义关系的概念集合,其中最典型的语义关系即为上下位关系.

**定义1** 上下位关系是一种基本的语义关系.给定概念  $c_1$  和  $c_2$ ,若  $c_2$  的外延包含  $c_1$  的外延,则认为  $c_1$  和  $c_2$  具有上下位关系.称  $c_2$  为  $c_1$  的上位概念,  $c_1$  为  $c_2$  的下位概念.

**定义2**<sup>[8]</sup> 概念空间是一个由概念及其之间的语义关系构成的集合.

概念空间用1个2元组  $S = (C, R)$  表示,其中称集合  $C$  为概念集合,称  $R \subseteq C \times C$  为上下位关系的集合,若  $c_1, c_2 \in C$ ,且  $(c_1, c_2) \in R$ ,则  $c_2$  为  $c_1$  的直接上位概念.又因为  $R$  具有传递性,通过两两概念之

间的上下位关系可合并为多层次树状结构,因此概念空间是1个多级结构.

**定义3** 在1个多级概念空间  $(C, R)$  中,若  $(c_1, c_2) \in R$ ,则称  $c_1 \rightarrow c_2$  为纵向推理,记为  $c_1(<) \mapsto c_2$ ,其可信度记为  $\theta_{c_2}(c_1) = P(c_2 | c_1)$ ,  $\theta_{c_1}(c_2) \in [0, 1]$ .相反,称  $c_2 \rightarrow c_1$  为回溯推理<sup>[9]</sup>,记为  $c_2(>) \mapsto c_1$ .

**定义4** 情感图式是使用图式形式表征能够唤醒某种情感的情境刺激和评价应付行为,其本身包含不同层次的概念,两两概念之间具有上下位关系,整体表现为1个多级结构.

在情感图式形式结构上,本文将诱发情感的情境刺激及可能的评价应付行为映射到多级概念空间上,以树形结构表示情感图式,树中每个节点代表1个刺激事件和对应的评价应付行为,边表示不同节点之间存在上下位层次关系.不同深度的节点代表不同层次等级,整体呈现多级性.自叶子节点向上,随着节点高度增加,概念的抽象化程度逐渐提高.

## 2 情感图式的构造及应用

### 2.1 情感图式构造

文献[5]详细描述了引起人类4种基本情感的情境刺激分类体系,包括快乐、悲伤、恐惧和愤怒.在这些分类体系的基础上,结合认知语境<sup>[10]</sup>及概念空间,本文构造了4种情感图式.构造流程如图1所示.

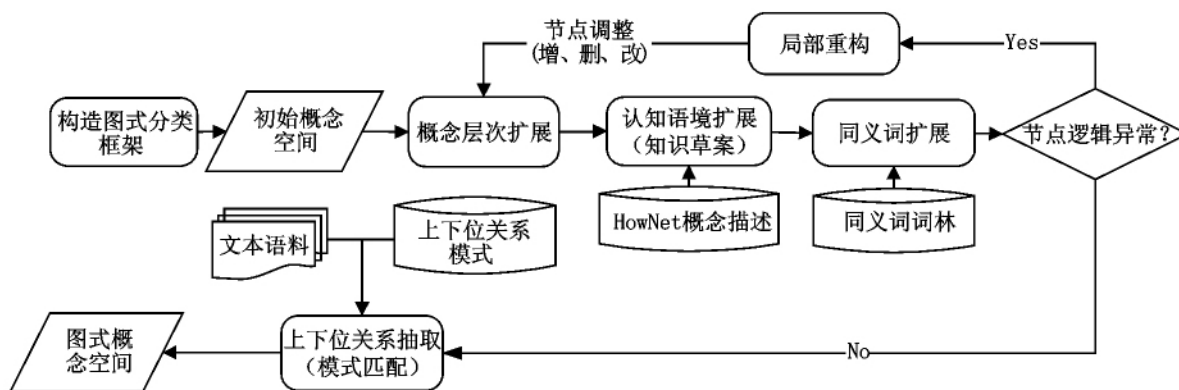
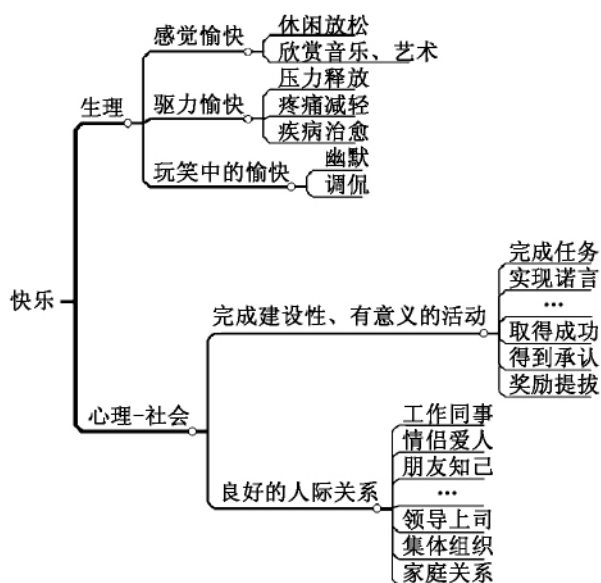


图1 情感图式构造流程图

**2.1.1 图式分类框架** 依据情绪心理学中对情境刺激的分类体系构造情感图式的基本框架,形成情感图式的初始概念空间.以“快乐”图式为例,快乐的诱因来自于生理满足和社会—心理方面的满足,其中生理上的满足分为感觉愉快、驱力愉快和玩笑中的愉快.社会—心理上的满足包含鲜明的社会内

涵,包括完成建设性、有意义的活动和良好的人际关系.前者包括完成任务、实现诺言、得到社会承认等;而后者包括朋友关系、同事关系、家庭关系等.据此描述构造的“快乐”图式框架如图2所示.同时,按照“恐惧”、“愤怒”、“悲伤”的情境刺激分类描述,本文又分别构造了3种情感的图式框架.



### 2.1.2 概念层次扩展、认知语境扩展及同义扩展

以人工标注方式结合百科知识进行概念层次扩展。采用由抽象到具体的构建过程,子节点是其上层父节点的具体化。为规范化扩展过程,提出了1个假设和1个构造原则:

(i) 假设: 每个情感图式最终构成的树中叶子节点不再具有下位概念,不可进一步扩展。

(ii) 构造原则: 从下位节点到上位节点的纵向推理过程可信度为1,即图式构建过程中,需保证不能上下位颠倒或在互不相干的概念之间构造上下位关系。

认知语境包括语言的情境知识、语言上下文知识和背景知识3个语用范畴,是语言使用者进行知识推导的主要依据。认知语境的操作以知识草案为基本单位,知识草案是基本行为单位的认知结构。语用推理是语用者运用一定的知识与策略,推导话语与语境的关联,并解读由此产生的语境暗含的思维过程。

本文使用 HowNet 中的概念义原构造知识草案,对非叶子节点进行词法分析并进行扩展,如“医院”(知识草案: DEF = InstitutePlace | 场所, @ cure | 医治, #disease | 疾病, medical | 医)。使用知识草案中的相关信息构造认知语境,通过实施者、接受者、场所等背景知识和认知语境信息进行语用推理。图式中叶子节点不再具有下位概念,可使用同义词林扩展,以增加图式的适用范围。如“晋升”的同义扩展为“提升,升级,晋级,···,升迁”。

由于词的多义性及扩展过程带来的噪音,概念

空间中概念的上下位关系不都符合构造的假设和原则。需要进行节点的删除或位置修改,通过局部重构,重新校正节点间的上下位关系,直到没有明显逻辑错误为止。

2.1.3 基于模式匹配的上下位关系自动获取 以手工标注为主的方法虽然保证了准确率,但限制了情感图式的规模。同时,由于标注者本身的知识和阅历局限性以及性格差异,图式的普适性不能保证。因此,需要使用一种机器自动抽取技术从文本集合中识别上下位关系词汇,有效地扩充图式。目前上下位关系获取方法主要有2种:基于模式的上下位关系获取方法<sup>[8]</sup>和基于统计的上下位关系获取方法<sup>[11]</sup>。前者主要利用语言学 and 自然语言处理技术,通过模式匹配获取上下位关系,该方法适合语法表达规范的语料;后者主要采用统计语言模型,需要有丰富的语料统计信息为基础。考虑到本文的语料来自于教材及期刊杂志,表达规范,使用基于模式方法效果更好。

结合汉语常用的表达方式,本文选用了5种上下位匹配模式<sup>[12-13]</sup>如表1所示,其中NP1表示下位名词短语,NP2表示上位名词短语。

表1 中文上下位关系模式

模式标签	中文上下位关系模式
P1	〈NP1〉〈是 为〉〈数词(/n)〉〈量词(/q)〉〈NP2〉
P2	〈NP1〉、···、〈NP1〉〈等等 等 和其他 以及其他 及其他〉〈NP2〉
P3	〈NP1〉〈是 为〉〈最 较〉〈/a〉〈的〉〈NP2〉
P4	〈NP1〉〈是 为〉〈NP2〉〈之一 的一(量词/q)〉
P5	〈NP2〉[·,]〈例如 如 比如 譬如 像 包括 有 特别是 尤其是〉〈NP1〉、···、〈NP1〉

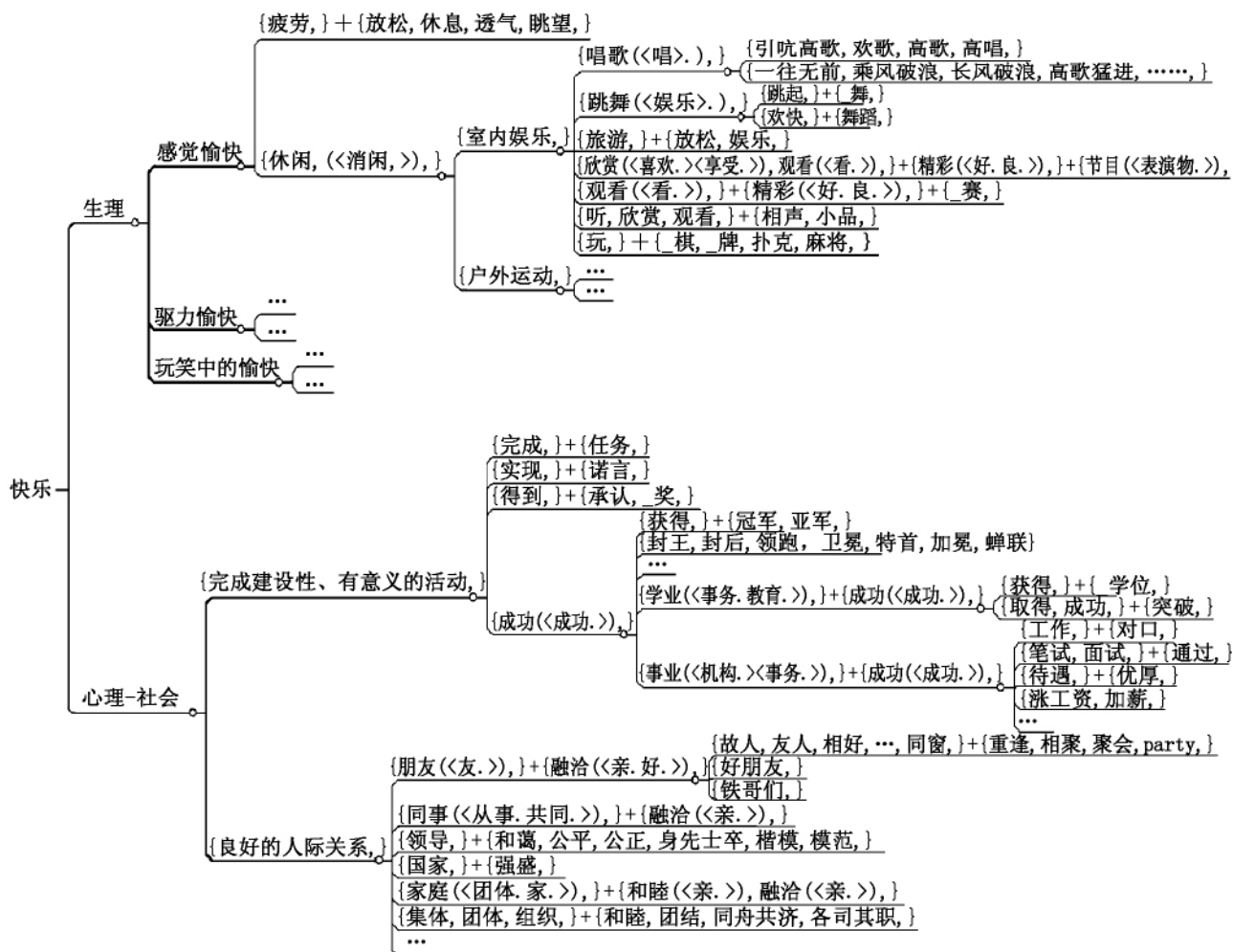
通过模式匹配,可从训练集中识别上下位关系词。对情感图式中节点进行进一步扩展,最终扩展后的快乐情感图式(局部)如图3所示。

### 2.2 基于语用推理和概念层次推理的情感认知模型

结合上文构造的情感图式,在徐琳宏等的情感认知模型的基础上,提出了基于语用推理和概念层次推理的情感认知模型( $E$ 代表当前句的情感)

$$E = (E_{\text{Prev}} E_{\text{Tran}} F_{\text{Seman}} C_{\text{Cong}}).$$

考虑到情绪发出者情绪具有一定的连贯性和持续性, $E_{\text{Prev}}$ 代表上一句的情感,作为当前句情感的一个重要的影响因子; $E_{\text{Tran}}$ 代表7大类情感间的情感迁移规律<sup>[14]</sup>,从情感语料库中统计转移概率较低的情感为有极性关系的情感; $F_{\text{Seman}}$ 代表语义特征,包



括情感词汇、否定词以及句间连接词等;  $C_{\text{Cong}}$  是在认知语境和推理理论的融合, 其中

$$C_{\text{Cong}} = (C_W, C_B, C_L).$$

该模型在综合多种不同的情感影响因子的同时, 有效融合了语用推理和概念层次推理理论, 增加了模型的适用范围和灵活性, 有助于提高文本情感识别的准确率。

### 2.3 情感图式在句子级文本情感计算中的应用

单纯使用情感词和语义特征进行的文本情感分析难以发现不含有明显情感词和情感语义特征的文本的情感; 此外, 含有显性情感词标识的句子, 有时情感词所传达的情感却不能代表句子的真实情感。因此, 需要借助上文的基于语用推理和概念层次推理的情感认知模型, 通过遍历每种情感的情感图式树, 进行推理匹配。推理过程中要用到上文中的纵向推理, 它是一种由下位概念推导上位概念的推理过程。若文本中没有出现当前概念空间中的情境刺激, 可以使用其上位概念进行判断。整个推理过程采取

自底向上的方式, 具体过程为

(i) 每个节点采取关键词匹配的方式, 匹配成功则转(iv), 否则转向(ii);

(ii) 结合语用推理知识, 使用该节点的知识草案分析判断, 若包含全部的概念义原, 则匹配成功, 转(iv), 否则转(iii);

(iii) 回溯至当前节点的父节点进行再匹配。重复(i)过程; 若回溯至根节点仍没有匹配则表明情感图式无法判断其情感, 需借助情感词和语义特征做进一步分析;

(iv) 输出情感类别信息, 退出程序。

本文的句子级情感分析方法使用了大连理工大学的情感词汇本体库, 在情感词的基础上, 使用本文构造的情感图式作为一种辅助的情感分析手段为每个句子建立1个6维向量, 每一维代表1种情感。每个句子对应1个情感词和语义特征得分向量  $v_{\text{lexicon}}$  和情感图式特征得分向量  $v_{\text{schema}} \cdot tf_{e_i}$ , 代表每个句子中第*i*类情感词的词频, 归一化后得到

$$v_{\text{lexicon}} = [tf_{e_1} \ tf_{e_2} \ tf_{e_3} \ tf_{e_4} \ tf_{e_5} \ tf_{e_6}] / \sqrt{\sum_{i=1}^6 tf_{e_i}^2}, \quad (3)$$

$$v_{\text{schema}} = [d_1 \ d_2 \ d_3 \ d_4 \ d_5 \ d_6] / \sqrt{\sum_{i=1}^6 d_i^2} \quad (d_i = \alpha_i \cdot s_i, \ s_i \in \{0, 1\}, \ d_i \in [0, 1]), \quad (1)$$

其中(1)式中 $s_i$ 代表句子中第 $i$ 类情感的对应的情感图式得分信息,该值的计算是通过情感图式进行推理得到的布尔值.考虑到 $s_i$ 的值只有0和1之分,在句子含有多种情感时区分度较小,故在每个维 $s_i$ 之前加入1个情感比例系数 $\alpha_i$ , $\alpha_i$ 是由语料中通过统计得到的每种情感句子数占总句子数的比例.经过向量归一化后得到向量 $v_{\text{schema}}$ .

采用线性融合方式将向量 $v_{\text{lexicon}}$ 和 $v_{\text{schema}}$ 融合,

$$v = t v_{\text{lexicon}} + (1 - t) v_{\text{schema}} \quad t \in [0, 1], \quad (2)$$

其中融合系数 $t$ 反映了情感词汇语义特征和情感图式特征两者在文本情感分析中的重要程度.

最终,每个句子的情感即为向量 $v$ 各维中得分最高的情感类别标签.

### 3 实验与结果分析

本文语料采用 DUTIR 的情感语料库<sup>[14]</sup>,选取其中的剧本、文学期刊、小学课文和童话部分,共计 990 篇文章,43 537 个句子,864 944 个单字.实验将语料平均分为 4 份,采用 4 倍交叉验证方式.实验结果评估方法为单句评估方法:以句子为单位,一类情感中判断正确的句子数量除以该类情感的句子总数.此外,还采用微平均来衡量情感分析的准确率.微平均是指所有标注正确的句子数除以总句子数.

首先,由于目前的情感计算方法大都以判断文本的倾向性为目的,只需判断文本褒贬而无需甄别其具体属于哪种情感类别;其次,文章论证的是借助情感图式为情感分析研究提供一个新角度,使其作为情感分析的一种辅助手段来提高情感分析效果.因此,选用以情感类别分析为目的且目前为止效果比较好的徐琳宏等的方法作为 Baseline.该方法使用的基于本体和语义特征的句子级文本情感分析方法,在情感词典的基础上,将影响句子情感色彩的 9 个语义特征加入条件随机域(CRF)中进行文本情感分析.本文的实验方案如下:

方案 1:单独使用情感词典<sup>[15]</sup>,采用关键字匹配的方法,计算句子中各类情感词的词频,词频最高的情感即为该句的情感.

方案 2:单独使用情感图式,结合语用推理和概

念层次推理进行句子情感分析.

方案 3:在方案 1 中分别单独加入悲伤、愤怒、恐惧、快乐情感图式,求得 4 组结果与方案 1 进行对比.

方案 4:在方案 1 中同时加入 4 种情感图式,使用上文提出的基于语用推理和概念层次推理的情感认知模型,与 Baseline 结果进行对比.

对于(2)式中的比例系数 $t$ ,本文以方案 4 为依托,以 0.1 为梯度,求得在不同梯度值下准确率的微平均值进行比较.由图 4 可知,当 $t=0.8$ 时,微平均达到 1 个峰值,故当 $t$ 取 0.8 时,情感分析效果较好.此时,实验结果如表 2 所示(情感类别中“平静”代表无情感).

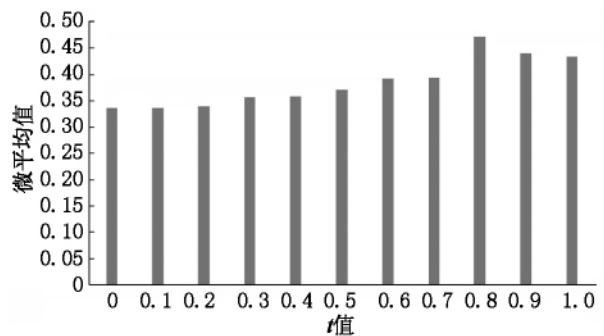


图 4  $t$  值与微平均值关系图

由表 2 可以看出,方案 3 在情感词典基础上加入单个情感图式,与方案 1 比较,加入图式的结果相比于没加之前有了 4%~7% 的提高,但微平均有所下降.这是由于加入情感图式后势必会带来不同程度的噪音词汇,其最大的影响为将本来无情感的句子误判为有情感的句子,而无情感句子在整个语料中占了将近 40% 的比例,它的准确率波动势必会造成整个微平均值的波动.

在方案 4 中加入了所有的情感图式,相比方案 1 微平均值有明显提升.“惧”的准确率下降是因为一种刺激可能引发多种情感,从而导致不同的情感图式之间有交集,当多种情感同时存在时,语料中句子数比例较大的情感占优势,比如“喜”和“哀”.对比方案 4 与 Baseline 结果,加入 4 种情感图式后,情感分析准确率都有较大的提高.但是情感图式的加入或多或少带来了噪音,导致部分本就无情感的句子误判为有情感;而“怒”类情感句子在语料中占有的比例较少(只有 2%),其准确率受噪音词的影响波动较大.从整体来看,微平均值相比 Baseline 提高了 4%,表明情感图式和基于语用推理和概念层次推理的情感认知模型可以有效地辅助传统方法进行文本情感分析,提高文本情感分析的准确率.

表 2 实验结果

情感类别	方案 1	方案 2	悲伤图式	愤怒图式	恐惧图式	快乐图式	所有图式	Baseline
哀	0.166 3	0.173 7	0.211 1	0.172 4	0.172 4	0.171 9	0.234 2	0.120 2
惊	0.149 0	—	0.152 0	0.152 0	0.152 0	0.151 9	0.169 7	0.089 5
平静	0.608 5	0.565 1	0.554 3	0.533 9	0.571 8	0.506 9	0.551 2	0.687 2
怒	0.069 6	0.076 9	0.072 3	0.135 5	0.072 3	0.070 2	0.292 6	0.474 4
恶	0.211 1	—	0.232 5	0.232 4	0.232 5	0.232 4	0.431 9	0.483 8
喜	0.579 5	0.391 4	0.586 9	0.586 8	0.586 9	0.648 4	0.577 6	0.353 4
惧	0.193 8	0.181 8	0.213 2	0.213 2	0.233 4	0.210 7	0.133 5	0.009 2
微平均	0.435 2	0.327 9	0.424 4	0.415 1	0.429 3	0.417 9	0.471 9	0.431 4

4 结束语

本文从认知学和情绪心理学角度出发,以图式理论及认知评价理论为依托,重新定义了情感图式。结合情绪心理学、语用推理和概念层次推理理论构造了 4 种情感图式,并在这 4 种情感图式基础上提出了一种改进的情感认知模型。将该模型应用于句子级文本情感计算中,实验结果表明情感图式能较好地辅助传统情感分析方法进行文本情感计算。

由于没有公开地为用于情感类别分析的其他语料,所以实验只能在现有文本比较规整的语料集上展开。情感图式在不规则文本上的泛化能力是未来研究工作的一个重点。构造情感图式中需要人工标注工作,采用机器学习方法,与统计方法相结合的方法也有待进一步研究。

5 参考文献

[1] R W Picard. Affective computing [M]. Cambridge MA: MIT Press ,1997.  
[2] Tao Jianhua ,Tan Tieniu. Affective computing: a review [M]. USA: Springer 2005: 981-995.  
[3] 徐琳宏,林鸿飞. 基于语义特征和本体的语篇情感计算 [J]. 计算机研究与发展 2007 44( S2): 356-360.  
[4] 徐琳宏,林鸿飞. 认知视角下的文本情感计算 [J]. 计

算机科学 2010 37( 12): 182-185.  
[5] 孟昭兰. 情绪心理学 [M]. 北京: 北京大学出版社, 2005.  
[6] Just Marcel Adam. The psychology of reading and language comprehension [M]. MA: Allyn & Bacon ,1987.  
[7] 熊学亮. 认知语用学概论 [M]. 上海: 上海外语教育出版社 ,1999.  
[8] 韩家新,何华灿. 基于抽象度的概念层次上的推理方法 [J]. 计算机工程与设计 2007 28( 3): 635-637.  
[9] Ekman P. Facial expression and emotion [J]. American Psychologist ,1993 48( 4): 384-392.  
[10] 徐琳宏,林鸿飞,潘宇,等. 情感词汇本体的构造 [J]. 情报学报 2008 27( 2): 180-185.  
[11] Yang Liang ,Lin Hongfei. Construction and application of chinese emotional corpus [C]. Berlin: Springer 2012: 122-125.  
[12] 刘磊,曹存根,张春霞,等. 概念空间中上下位关系的意义识别研究 [J]. 计算机学报 ,2009 ,32( 8): 1651-1661.  
[13] Tian Fang ,Ren Fuji. Hyponymy acquisition from Chinese text by SVM [C]. Tokushima: Univ of tokushima press , 24-27.  
[14] Sharon A. Caraballo. Automatic construction of a hypernym-labeled noun hierarchy from text [C]. Maryland: Brown Univ press ,1999: 120-126.  
[15] Jiang Peilin ,Xiang Hua ,Ren Fuji et al. The Framework of Mental StateTransition Analysis [C]. Mexico: Springer , 2007: 1046-1055.

The Construction of Affective Schemata and Its Application  
in Text Affective Computing

REN Ju-wei ,YANG Liang ,LIN Hong-fei\*

( Information Retrieval Laboratory ,Dalian University of Technology ,Dalian Liaoning 116024 ,China)

**Abstract:** Stimulus situation has been mapped to multi-level concept space on the basis of cognitive evaluation theory. With HowNet and synonyms dictionary ,the method of hyponymy acquisition based on patterns to construct four kinds of affective schemata has been used. Combined with the cognitive context and pragmatic inference ,these affective schemata have been used in sentence level text affective computing. The experiments show that schema can effectively assist the recognition of text affection and improved its precision.

**Key words:** affection schemata; affective computing; concept space; cognitive context; pragmatic inference

( 责任编辑: 冉小晓)