

汉语情感词的极性： 基于新加坡华语使用者的情感词极性词库

Chin Loong Ng¹; Jingxia Lin¹; Yao Yao²

¹Nanyang Technological University; ²Hong Kong Polytechnic University

CNG013@e.ntu.edu.sg; jingxialin@ntu.edu.sg; ctyaoyao@polyu.edu.hk

摘要：本文以研究汉语词汇的情感极性（polarity）为主，采集多位非专业人士（即普通的汉语使用者）对汉语词语的极性判断，并对此进行数据统计获得汉语词汇的极性均值。此外，本研究还讨论了情感极性与情感类别及情感强度之间的相关性，并与前人的极性研究结果进行了比较分析。本研究的调查对象仅限于新加坡华语使用者，然而，本文所呈现的研究方法和结果对将来更为大型的汉语情感词极性研究或相关的情感研究都具有重大参考意义与借鉴作用。

关键词：情感词，情感极性，新加坡华语

Polarity of Chinese Emotion Words: The Construction of a Polarity Database Based on Singapore Chinese Speakers

Abstract: In this paper, we report a study of the polarity of Chinese emotion words. We conducted a large-scale polarity rating experiment with laymen speakers, and compiled a database of polarity ratings for Chinese emotion words based on these experimental results. The polarity ratings were also compared with previously reported polarity ratings, as well as related emotion word ratings such as emotion category and emotional intensity. The participants in the current study were all Singapore Chinese speakers, but the methodology and the current results will serve as an important reference for future research on sentiment analysis and emotion language in Chinese in a broader context.

Keywords: emotion word, emotion polarity, Singapore Chinese.

1 前言及文献综述

文本情感分析（sentiment analysis）属于意见挖掘（opinion mining; Ding et al. 2008; Esuli & Sebastiani, 2006, etc.）一类的研究，采用计算语言学的方法来计算和分析文本的情感倾向。情感极性分析是文本情感分析的一种，主要目的是判断文本所表达的情绪倾向及其强度（Esuli & Sebastiani, 2005; Wilson, Wiebe, & Hwa, 2004, etc.）。情绪倾向可以是正面的（positive），负面的（negative），也可以是中性的（neutral）；情感极性强度，即情绪倾向的强度，则往往用数量或等级来表示。蕴含情感倾向的词可以称为极性词，也可以称为情感词。在本文中，这两个名称可以通用。

抽取及分析极性词目前主要有两种研究方法，基于语料库，或是基于词典。基于语料库的代表作有 Hatzivassiloglou & McKeown (1997) 和 Wiebe (2000) 等。Hatzivassiloglou & McKeown (1997) 指出当一个句子中出现连词连接形容词，可依据该连词的语法作用和首个形容词的正负属性来推测出后方形容词的正负属性。他们从巨型语料库中提取类似句子，通过线性回归模型来分析形容词的情感极性，再以聚类算法将形容词分门别类后标上该词的情感极性。Wiebe (2000) 首先以人工方式标注种子形容词的情感极性，接着从语料库中统计其他形容词与种子词的互信息后，更精准地判断新词的情感极性。Turney & Littman (2003) 采用两种统计方法来判断词的情感倾向，即点互信息（Point Mutual Information）以及潜在语义分析（Latent Semantic Analysis），研究对象包括形容词、副词、名词和动词。Ding et al. (2008) 提出特定领域情感词的配对研究方法（holistic lexicon-based approach），这个方法让统计系统更精准地计算词汇在特定情境中所蕴含

的情感极性。此外，Ding 等也将语境纳入考量，将会影响情感判断的特定词汇、短语和语法结构等引入 Opinion Observer 统计系统中。

第二类基于词典的代表作包括 Kamps et al. (2004) 和 Esuli & Sebastian (2005)。Kamps et al. (2004) 以图论法 (graph theoretic model) 着重研究 WordNet 中极为关键的同义性，以基于语义词典的方法来构建一个类似 WordNet 形容词极性词库。Esuli & Sebastian (2005) 使用定量分析 (quantitative analysis) 方法来判定词汇的情感倾向。研究利用网络词典的同义词来扩大种子词集，然后从词典中提取词的释义作为文本分类基准。扩大后的种子词集和词典释义将会用来训练文本分类器，之后由该分类器根据已有的释义信息来判断新词汇的情感极性。Andreevskaia & Bergler (2006) 推出一项以词典释义为基础的系统 (Sentiment Tag Extraction Program)，利用 WordNet 中的释义来识别并判断词汇的情感极性。Kim & Hovy (2005) 利用词典将人工采集的种子情感词集进行扩增。为避免搜集冷门词汇，采取让外国学生将该基础英语词汇表进行 TOEFL 英语测试。之后随机抽取 462 形容词与 502 动词，并让 3 名标注者将词汇标注为极性词或非极性词。此外，也有学者从句子和篇章的篇幅来进行情感极性分析。当然，也有研究兼用了基于词典和基于语料库两种研究方法，比如徐等 (2008)。

综上所述，在前人研究中，往往只有很少的人员参与到极性词标注的过程中。极性的标注要么是机器自动进行的，要么是通过少量的标注者 (多数是有经验的研究人员) 完成的。显然，这样的标注过程虽然高效，但却未必可以完全反映出普通的语言使用者对于文本中蕴含的情感极性的感知。针对这个问题，在本次研究中，我们邀请大量的普通语言使用者参与到情感极性标注的过程。

2 研究方法

本调查所采用的研究语料和研究方法和 Lin & Yao (2015) 非常相似。Lin & Yao (2015) 介绍了一个在中国大陆进行的关于汉语情感词的情感类别和情感强度的研究。Lin & Yao (2015) 的词表筛选自许和陶 (2003)，研究方法则来源于 Strauss & Allen (2008)。在本研究中，我们完全沿用了 Lin & Yao (2015) 的词表，并大量参考了他们的研究方法。相对于前人的情感词极性研究 (徐等 2008; Khoo et al., 2015; etc.) 而言，本研究所要调查的词语数目较少，且主要以情感词为主，但作为第一个基于大量华语使用者标注的汉语情感极性研究，本研究的方法和结果都可为将来更为大型的情感极性研究提供参考作用。

这项研究共涉及 372 个情感词，组成每个词的字数从 1 到 4 字不等。这些情感词被进一步划分为四份词表 (词表一到词表四)，每份词表包含 93 个词，每位受试者只需要完成其中一份词表。与 Lin & Yao (2015) 一样，本研究也采用了一定的控制举措以确保调查结果的一致性与可信度。首先，每份词表中都有两个情感词会在该词表中出现两次，以侦测每位受试者答题的一致性。一共有八个这样的“表内重复词”。其次，每两个词表都会有两个重叠的词，一共有六个这样的“跨表重复词”，用来测试各受试群答题的一致性。最后，每份词表都包含一个现代汉语中不存在的词语，即假词“几几”，该词也用作检测受试者是否认真作答。最终，每份词表将包含 97 或 98 道题。表 1 列出了所有的表内重复词和跨表重复词。

表 1 表内重复词和跨表重复词

表内重复词	跨表重复词
词表 1: 厌倦, 悲痛	词表 1, 2: 开心
词表 2: 担忧, 丧气	词表 1, 3: 焦躁
词表 3: 愉悦, 痛恨	词表 2, 4: 愤怒
词表 4: 愤慨, 愁闷	词表 2, 3: 沉痛
	词表 3, 4: 震惊
	词表 1, 4: 畏惧
	出现在所有词表的假词: 几几

这项研究的受试群主要来自在新加坡生活和接受教育的华语使用者。这项问卷调查共有 94 (男 35 女 59) 位受试者参与，年龄介于 16 到 30 岁，现今主要居住地在新加坡。所有受试者

都在新加坡完成基本的华语教育，他们至少完成了初级学院（Junior College）或理工学院（Polytechnic）程度以上的教育。

这项研究的问卷在谷歌表格界面（Google Form Platform）发布。受试者被随机分配完成一份词表的问卷。在问卷中，受试者被要求按照自己的语言直觉为词表中的每一个情感词标注情感极性。情感极性共有七个选项，分别是 3、2、1、0、-1、-2、-3。其中，选项“3”表示“极为正面/积极”，选项“0”表示“无正负色彩/中立”，选项“-3”表示“极为负面/消极”。另外，若受试者不明白该情感词的词义，则可以选择“不理解词义”的选项。

3 研究结果

本研究共由 89 名受试者参与标注，每个词表的受试者在 23-23 人之间。如前文所述，为了检验标注的一致性和可靠性，我们用了八个“表内重复词”和六个“跨表重复词”。我们用 t 测试（t test）检验所有重复词的两次标注。结果显示所有重复词的两次标注都没有显著差异（所有 p 值大于 0.1）。此外，对于假词“几几”的标注，一共有 92.1% 的标注者将“几几”标注为“中性词”或“不理解词义”。虽然仍有 7.5% 的标注者为假词“几几”标注上情感极性强度，但我们认为这并不影响本调查结果的可信度。首先，尽管“几几”在现代汉语中不是真正的词语，但“几”这个字却蕴含了多重语义，例如“几个”、“几何”等。其次，“几几”与汉语中的“唧唧”、“岌岌”、“叽叽喳喳”等存在谐音，也可能因此影响标注者的判断。另外，由于流行科技文化如网络和数码媒体经常会有产生新词的现象，因此一些标注者也可能将“几几”视为一个新兴的流行词汇。综合所有的测试结果，我们认为本实验获得的标注数据具有较高的可信度。

为方便分析和讨论，本研究在统计情感极性强度均值（Mean Polarity）之后，将这 372 个词汇根据情感极性强度均值来划分为三大类：偏负面极性词（ $-1 \geq \text{Mean Polarity} \leq -3$ ），偏中性极性词（ $-1 > \text{Mean Polarity} < 1$ ）以及偏正面极性词（ $1 \geq \text{Mean Polarity} \leq 3$ ）（当然，本文所采纳的仅是众多分法当中的一类，研究者可以依据研究所需，对情感极性强度均值进行不同的分类，比如将均值大于 0 的归类为“正面极性词”，小于 0 的归类为“负面极性词”）。整体而言，这 372 个词汇当中有 143 个偏负面极性词，占 38.4%，偏中性极性词有 125 个，占 33.6%，至于偏正面极性词则有 104 个，占 28.0%。在 125 个偏向中性极性词中，“挂牵”是唯一一个情感极性强度均值为 0 的词汇。偏负面极性词比偏正面极性词多出 10.4% 的调查结果与其他相关的情感研究有相似之处，比如 Khoo et al., (2015) 在比较英语情感词库时发现这些情感词库拥有一个共同的特点，那就是负面词汇的数量比正面词汇的数量来得多。表 2 所示的是按照本文的分类方法所划分出的三大类情感极性类别、词汇以及词数。在表 2 中，我们特别标记了极为正面的词（ $2 \geq \text{Mean Polarity} \leq 3$ ）和极为负面的词（ $-2 \geq \text{Mean Polarity} \leq -3$ ）。一共有 12 个极为负面的词以及 10 个极为正面的词。这些词语可以认为是最强烈最毋庸置疑的情感正（负）级。

除了情感极性强度均值以外，我们也计算了每一个词的极性强度范围，即最高极性强度标注和最低极性强度标注之间的差值。例如“骇异”最低标注为-2，最高标注为 0，差值为 2，而“快活”最低标注为-3，最高标注为 3，差值为 6。在所有的 372 个词当中，最小的差值是 2，而最大的差值为 6。令我们惊讶的是，大部分的情感词的标注差值大于 3，即标注值分布在 0 的两边。其中，21.5%（n=80）的词差值是 4，36.3%（n=135）的词差值是 5，12.9%（n=48）的词差值是 6。换句话说，总共有 70.7% 的词有至少两个标注者对于情感倾向表达了相反的意见。我们推测，导致巨大差值的原因或许可能是基于个体标注者对个别词语理解上的差异所造成的。例如 Khoo et al., (2015) 就曾指出，同一个词“torrid”在某语境中可以是正面理解，然而如果放置到另一个语境中却可能含有负面理解，因此不同标注者在为该词做标注时，出现较大的差异。这也从侧面证明了用多个标注者的必要性。

表2 情感词极性分类词表（极为正面或负面的词有下划线标记）

情感极性类别	词汇
偏正面	<p>乐于、拥护、放松、炽热、神往、在乎、快活、安宁、感动、赞同、可意、怡和、关切、晓畅、来劲、欢娱、称意、信服、放心、欢、狂热、称心、带劲、康乐、安心、爱好、舒坦、关心、高亢、如意、愿意、相信、得志、舒畅、松快、乐意、动心、欢欣、关怀、赏识、顺心、珍视、振奋、充实、喜欢、宽心、惊喜、体贴、欢悦、高昂、景仰、崇尚、崇拜、感激、满意、爽心、感谢、喜爱、窝心、欢喜、激昂、乐、欣慰、痛快、推崇、欢快、自信、欢乐、钟爱、信赖、快乐、佩服、畅快、开心、赞赏、欢愉、崇奉、自在、信任、钦佩、欣喜、高兴、尊敬、珍爱、喜悦、尊重、珍惜、舒服、器重、兴奋、愉悦、舒心、敬佩、敬慕、<u>孝敬</u>、<u>自豪</u>、<u>尊崇</u>、<u>景慕</u>、<u>敬仰</u>、<u>欢畅</u>、<u>崇敬</u>、<u>幸福</u>、<u>敬重</u>、<u>热爱</u></p>
偏中性	<p>为难、烦心、生气、遗憾、低沉、苦恼、愕然、熬心、窘、不满、困惑、头疼、感伤、操心、烦恼、苦闷、心虚、发慌、惊慌、抱愧、窝火、惊疑、不好过、苦、心慌、怀疑、惊愕、担心、焦急、羞怯、心急、不平、失望、惊诧、无奈、疑心、骇怪、害怕、寂寞、眼红、伤心、疯狂、吝惜、吓人、挂虑、紧张、多情、骄矜、诧异、悲壮、担忧、对不住、解恨、怕、慌张、火、看不惯、娇宠、炫耀、心浮、可惜、害臊、痴迷、不过意、憋气、吃惊、震惊、过敏、对不起、狂、无聊、消魂、迟疑、急、怜惜、疑惑、着急、不好意思、挂心、痒痒、缠绵、惊讶、抱歉、沉静、羞涩、奇怪、挂牵、动摇、感慨、牵挂、偏爱、害羞、眷恋、怜悯、挂念、自满、骄傲、无辜、激动、同情、留神、了解、想、心切、关注、惊奇、亢奋、惦念、怀念、得意、情愿、自爱、可心、闲适、投入、闲雅、倚重、理解、逍遥、高涨、快慰、炽烈、瞧得起、羡慕、体谅</p>
偏负面	<p><u>嫉恨</u>、<u>憎恶</u>、<u>痛恨</u>、<u>愤恨</u>、<u>忌恨</u>、<u>鄙夷</u>、<u>暴怒</u>、<u>窝囊</u>、<u>绝望</u>、<u>忿恨</u>、<u>悲恸</u>、<u>轻蔑</u>、怨恨、悲凄、颓丧、沮丧、悲愤、歧视、郁悒、鄙视、厌恶、悔恨、抑郁、沉痛、惨痛、哀痛、忌妒、惊恐、苍凉、哀怨、瞧不起、懊丧、愤慨、悲凉、阴郁、揪心、腻烦、惊惧、骄慢、沉郁、忿怒、懊悔、妒忌、心寒、悲切、自卑、悲哀、悲痛、蔑视、愤怒、悲枪、浮躁、愤懑、冤、忧郁、恐慌、愁闷、敌视、丧气、冤枉、怅惘、反感、惶恐、痛心、发憷、背悔、忧愁、烦躁、恼恨、消沉、惊骇、惶惑、愁苦、焦躁、厌烦、憋闷、悲愁、犯愁、哀戚、困窘、抱委屈、烦杂、鄙薄、恐惧、负疚、焦炙、颓唐、嫉妒、忧虑、伤感、自负、发愁、焦虑、自大、畏怯、烦躁、狂乱、惭愧、慌乱、羞惭、骄横、心酸、灰心、酸辛、厌倦、后悔、溺爱、不快、哀愁、妒嫉、懊恼、讨厌、辛酸、惆怅、糟心、委屈、冲动、捣乱、烦、烦乱、烦闷、畏惧、自傲、乏味、怠慢、焦渴、低落、反对、沉重、沉闷、激愤、烦人、不安、不悦、哀伤、心焦、悲伤、惊惶、愁、窘促、羞愧、郁闷、骇异</p>

4 讨论

我们将本次研究的结果和徐等（2008）对中文情感词汇的极性标注进行了比较。徐等（2008）一共标注了四类情感极性，即中性、褒义、贬义以及兼有褒贬两性。需要指出的是，徐等（2008）的研究结合了人工标注和自动运算的结果，涵盖了极大量的中文词汇，其中仅有 327 个词汇和本研究覆盖的词汇重合。这 327 个词汇在徐等（2008）中被分别标注为褒义词（ $n = 76$ ），中性词（ $n = 186$ ），贬义词（ $n = 64$ ），以及兼褒贬（ $n = 1$ ），但是其极性标注与本研究的结果差异较大。在徐等（2008）标注的 76 个褒义极性词中，只有 10 个（13.2%）词汇在本研究中标注为偏正面词，其余 66 个（86.8%）词汇均在本研究中标注为偏中性词。特别值得注意的是“悲壮”以及“沉静”二词，它们的情感极性均值均是负数（分别为 -0.62 和 -0.09）。同样地，徐等（2008）的 64 个贬义词当中有 58 个（91%）词汇在本研究中被归类为偏中性词，有 2 个（3%）词汇归类为偏正面词，只有 4 个（6%）词汇则归类为偏负面词。徐等（2008）的 186 个中性词只有 72 个（39%）在本文中也归类为偏中性词，另有 87 个（47%）词汇归类为偏负面词，27 个（15%）归类为偏正面词。最后，徐等（2008）中的兼褒贬词，“挂虑”，在本研究被归为偏中性词，情感极性均值为-0.71。至于“挂虑”究竟是中性还是在不同语境之下可褒可贬，尚留待学者深入研究。

我们认为，导致本次研究和徐等（2008）有较大差异的因素至少有两个。首先，徐等（2008）的研究对象是中国大陆使用的普通话而本调查则针对新加坡华语，二者属于不同区域的汉语语体，因此一些情感词在语义上可能存在差异。第二个可能因素是标注者及标注过程的差异。徐等（2008）采纳人工标注和自动计算结合的模式来获取数据，而本调查研究的数据则完全基于人工标注。另外，本研究的标注完全来自于普通语言使用者，而非专业研究人员，因此标注人员对于情感词的理解也可能有所差异。这些差异表明我们还需要进一步研究比较各种不同研究方法的有效性和准确性。

最后，我们也将本次极性标注的结果和 Lin & Yao（2015）所报告的情感分类和情感强度作了对比。我们发现，情感分类和情感极性存在着明显的关联。如表 3 所示，所有的“憎”类的词都被标注为偏负面的词。大多数的“怒”、“焦”、“惧”、“悲”类词也被标注为偏负面，但仍有少量属于偏中性的词。“喜”类词基本都被标注为偏正面的词。“惊”类词则基本被标注为偏中性，只有一个被标注为偏负面的词。情感强度和情感极性也存在着一定的关联性。我们对本次研究中获得的极性标注与 Lin & Yao 获得的情感强度标注进行皮尔逊积矩相关系数的运算，发现在偏负面词和偏正面词中，情感极性与情感强度之间的相关系数值较高（分别为 0.57, 0.35），而偏中性词的情感极性与情感强度之间的相关系数值则较低（0.16）。也就是说，对于真正有极性的情感词，普通语言使用者对于其情感极性的判断和对于其情感强度的判断是密切相关的。但是对于本身情感极性趋于中性的词，语言使用者对于其情感极性的判断和对于其情感强度的判断较为独立。

表 3 各情感类别的情感极性分布

Lin & Yao (2015) 中的 情感类别	本文调查结果			主要极性类别
	偏负面	偏中性	偏正面	
怒	12 (66.7%)	6 (33.3%)	0 (0%)	66.7% 偏负面
焦	19 (52.8%)	17 (47.2%)	0 (0%)	52.8% 偏负面
憎	21 (100%)	0 (0%)	0 (0%)	全部偏负面
惧	8 (66.7%)	4 (33.3%)	0 (0%)	66.7% 偏负面
喜	0 (0%)	4 (7.1%)	52 (92.9%)	92.9% 偏正面
悲	52 (78.8%)	14 (21.2%)	0 (0%)	78.8% 偏负面
惊	1 (9.1%)	10 (90.9%)	0 (0%)	90.9% 偏中性

5 结论

在本文中，我们介绍了一个在新加坡进行的采用普通汉语使用者标注的情感词极性研究。我们从标注结果中获得了一个新加坡华语的情感词极性的数据库，也显示了和前人所做的相关研究的联系与不同。我们希望，本研究能为将来更大规模、更全面的汉语情感语言的研究提供一个有益的参考。

参 考 文 献

- Andreevskaia, A., & Bergler, S. (2006). Mining WordNet for a Fuzzy Sentiment: Sentiment Tag Extraction from WordNet Glosses. In *EACL* (Vol. 6, pp. 209–216).
- Ding, X., Liu, B., & Yu, P. S. (2008). A holistic lexicon-based approach to opinion mining. In *Proceedings of the 2008 International Conference on Web Search and Data Mining* (pp. 231–240). ACM.
- Esuli, A., & Sebastiani, F. (2005). Determining the semantic orientation of terms through gloss classification. In *Proceedings of the 14th ACM International Conference on Information and Knowledge Management* (pp. 617–624). ACM.
- Hatzivassiloglou, V., & McKeown, K. R. (1997). Predicting the semantic orientation of adjectives. In *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics and 8th Conference of the European Chapter of the Association for Computational Linguistics* (pp. 174–181). Association for Computational Linguistics.
- Kamps, J., Marx, M., Mokken, R. J., & De Rijke, M. (2004). Using WordNet to Measure Semantic Orientations of Adjectives. In *LREC* (Vol. 4, pp. 1115–1118).
- Khoo, C., Cheon, N., & Basha, J. (2015). Comparison of lexical resources for sentiment analysis. Presentation at the *Singapore Symposium on Sentiment Analysis*. Singapore.
- Kim, S. M., & Hovy, E. (2005). Automatic detection of opinion bearing words and sentences. In *Companion Volume to the Proceedings of the International Joint Conference on Natural Language Processing (IJCNLP)* (pp. 61–66).
- Lin, J., & Yao Y. (2015). Encoding emotion in Chinese: Emotion type and intensity of Chinese emotion words. In *Proceedings of the 16th Chinese Lexical Semantics Workshop*.
- Strauss, G. P., & Allen, D. N. (2008). Emotional intensity and categorisation ratings for emotional and nonemotional words. *Cognition and Emotion*, 22(1), 114–133.
- Turney, P. D., & Littman, M. L. (2003). Measuring praise and criticism: Inference of semantic orientation from association. *ACM Transactions on Information Systems (TOIS)*, 21(4), 315–346.
- Wiebe, J. (2000). Learning subjective adjectives from corpora. In *AAAI/IAAI* (pp. 735–740).
- Wilson, T., Wiebe, J., & Hwa, R. (2004). Just how mad are you? Finding strong and weak opinion clauses. In *AAAI* (Vol. 4, pp. 761–769).
- 徐琳宏 等. (2008) 情感词汇本体的构造. 《情报学报》, 27 (2), 180–185.
- 许小颖, 陶建华. (2003). 汉语情感系统中情感划分的研究. 第一届中国情感计算及智能交互学术会议论文集, 199–205.