

# 基于语法的情感词汇自动获取

陈建美,林鸿飞,杨志豪

(大连理工大学 电子与信息工程学院,辽宁 大连 116024)

**摘 要:**情感计算是目前人工智能领域的热门课题,而词汇的情感计算又是准确完成文本情感计算的基础.目前情感词汇的获取大多采用人工获取的方法,如何自动地获取情感词汇,已成为当前情感计算研究亟需解决的问题.提出了情感词汇的自动提取机制,首先分析了情感词汇的一般语法规律,例如,重叠的规律,受否定词、程度副词修饰的规律等.然后在情感词汇的这些语法规律的基础上,运用 CRF 模型实现了情感词汇的自动获取.最后,分析了不同的语法规律对情感词汇自动获取的作用大小,并对实验结果进行了详细分析,实验结果表明情感词汇自动获取方法是有效的.

**关键词:**情感词汇;词汇自动获取;情感计算;条件随机域

**中图分类号:** TP391.3 **文献标识码:** A **文章编号:** 1673-4785 (2009) 02-0100-07

## Automatic acquisition of emotional vocabulary based on syntax

CHEN Jian-mei, L N Hong-fei, YANG Zhi-hao

(School of Electronic and Information Engineering, Dalian University of Technology, Dalian 116024, China)

**Abstract:** Affective computation has received more and more attentions in the field of artificial intelligence; however, the calculation of affective lexicon ontology is a requirement for affective computation of texts. At present, most emotional lexicons are obtained by manual methods. The automatic acquisition of emotional lexicons has become an urgent task that needs to be addressed. This paper presents an automatic acquisition method for emotional lexicons. The authors analyzed the general syntactical rules of emotional lexicons, such as the rules of overlapping words, then rules governing how these acquired emotional words were modified by privatives and degree adverbs. Then we used the conditional random fields (CRF) model to acquire emotional words based on the general rules. Finally, we analyzed the effects of various syntax rules on the automatic acquisition of emotional vocabulary. Experiments were done and the results showed that the proposed method is effective for automatic acquisition of emotional words.

**Keywords:** emotional vocabulary; automatic vocabulary acquisition; affective computer; conditional random field

人类之间的沟通与交流是自然而富有感情的,因此,在人机交互的过程中,人们也很自然地期望计算机具有情感能力.情感计算<sup>[1]</sup>就是要赋予计算机类似于人一样的观察、理解和生成各种情感特征的能力,最终使计算机像人一样能进行自然、亲切和生动地交互.目前在情感计算方面所做的研究主要集中在声音、图像、生理信号和文本几方面.随着 Internet 的发展,以文本形式出现的信息越来越多,逐渐成为最容易获取也是最为丰富的一种交互资源.计

算文本的情感在许多方面有着广阔的应用前景,如颜色条表示文章的情感结构<sup>[2]</sup>、自动聊天系统<sup>[3]</sup>、文语转换、舆论监控等等.

词汇的情感计算是文本的情感计算的基础.目前,关于情感词汇方面的资源比较少,英文的有 WordNet,可以利用 WordNet Domain 对其进行分类,划分出 WordNet-Affect 类.中文的有知网的“情感分析用词语集”<sup>[4]</sup>,该词语集将情感分为正面和负面 2 种.但是人类的情感却不是非正即负的,而是喜、怒、哀、乐等丰富多彩的.因此,细致而准确的文本情感计算,需要多类别的情感词汇的支持.

收稿日期: 2008-12-16

基金项目: 国家自然科学基金资助项目 (60373095, 60673039); 国家“863 计划”资助项目 (2006AA01Z151); 教育部留学人员回国启动基金资助项目 (教外留司 [2007]1108 号).

通信作者: 林鸿飞. E-mail: hflin@dlut.edu.cn

## 1 情感词汇本体库

大连理工大学的情感词汇本体将情感分为 7 大类 20 个小类<sup>[5]</sup>,如表 1 所示.情感词汇本体通过一个三元组来描述:

$$\text{Lexicon} = (B, R, E).$$

式中: $B$ 表示词汇的基本信息,主要包括编号、词条、对应英文、词性、录入者和版本信息; $R$ 代表词汇之间的同义关系,即表示该词汇与哪些词汇有同义的

关系; $E$ 代表词汇的情感信息,包括情感类别、情感强度、情感极性,是情感词汇描述框架中比较重要的一部分.

情感本体的基本知识主要来源于现有的一些词典和语义网络.其中词典包括《现代汉语分类词典》<sup>[6]</sup>、《汉语褒贬义词语用法词典》<sup>[7]</sup>、《中华成语大词典》<sup>[8]</sup>等等.语义知识网络有知网和 WordNet 另外还加入了《汉语情感系统中情感划分的研究》<sup>[9]</sup>中的部分词汇.

表 1 情感分类  
Table 1 Affective classification

编 号	情感大类	情感小类	例 词	数量 /个
1	乐	快乐	喜悦、欢喜、笑咪咪、欢天喜地	840
2		安心	踏实、宽心、定心丸、问心无愧	234
3		好	尊敬、恭敬、敬爱、毕恭毕敬、肃然起敬	451
4		赞扬	英俊、优秀、通情达理、实事求是	5 602
5	怒	相信	信任、信赖、可靠、毋庸置疑	204
6		喜爱	倾慕、宝贝、一见钟情、爱不释手	765
7		愤怒	气愤、恼火、大发雷霆、七窍生烟	264
8		悲伤	忧伤、悲苦、心如刀割、悲痛欲绝	659
9	哀	失望	憾事、绝望、灰心丧气、心灰意冷	243
10		疚	内疚、忏悔、过意不去、问心有愧	85
11		思	相思、思念、牵肠挂肚、朝思暮想	127
12		慌	慌张、心慌、不知所措、手忙脚乱	202
13	惧	恐惧	胆怯、害怕、担惊受怕、胆颤心惊	338
14		羞	害羞、害臊、面红耳赤、无地自容	74
15		烦闷	憋闷、烦躁、心烦意乱、自寻烦恼	712
16		憎恶	反感、可耻、恨之入骨、深恶痛绝	1 086
17	恶	贬责	呆板、虚荣、杂乱无章、心狠手辣	5 025
18		妒忌	眼红、吃醋、醋坛子、嫉贤妒能	44
19		怀疑	多心、生疑、将信将疑、疑神疑鬼	67
20		惊奇	奇怪、奇迹、大吃一惊、瞠目结舌	134

目前,该情感词汇本体已收录情感词汇 17 156 个,为段落级和文本级的情感计算提供了基础和依据.但是,这 17 156 个情感词汇的获取基本采用人工的方法,既费时又费力,如何在现有的情感词汇资源的基础之上自动地获取情感词汇,从而节省人工成本,成为了亟需解决的问题.

## 2 相似研究

关于情感词汇的自动获取,目前还没有相关的

研究.与之相似的研究有:领域词汇的自动获取和倾向性词汇的自动获取.领域词汇的自动获取,目前的方法主要有:前景背景语料的方法<sup>[10]</sup>、领域词汇特定模式匹配的方法<sup>[11]</sup>、以及计算未知词汇和种子词汇 PMI(pointwise mutual information)的方法<sup>[12]</sup>.倾向性词汇的自动获取,目前的方法主要有模式匹配<sup>[13]</sup>和计算 PMI<sup>[14-15]</sup>.

文献[10]中使用前景背景语料的方法,来获取体育新闻领域的词汇.前景背景语料的构建需要有

明显的领域界限,而情感词汇没有明显的领域界限,即使是在相对而言比较客观的说明文中,也常常包含情感词汇。

文献[13]中使用“and”、“or”、“but”作为模式,来获取有倾向性的形容词对。这种方法仅仅考虑了形容词,对于动词、名词、副词等,则不予考虑。而且模式有限,难以扩充。

文献[14]认为汉语词汇的少量 Morpheme(词素)可以很好地指示大量汉语词汇的语义倾向,选取含有强烈语义倾向的 Morpheme,如:“奖、胜、优、富、健、欢、荣、努、顺”,“伤、贪、疑、困、急、妄、禁、拒”,使用 PMI 的方法,进行词汇的自动语义倾向判断。但是文献[14]的作者认为,这种方法只在小规模的语料范围内是有效的。

文献[15]中使用 AltaVista 搜索引擎索引的大规模语料,选用具有强烈语义倾向的词汇,如“excellent”、“poor”作为基准词,计算 PMI 来判断词汇的语义倾向。

倾向性词汇是指褒贬二义的词汇,而情感词汇还包括较多的中性词汇,例如表1中的“大吃一惊”不是一个倾向性的词汇,但是表达“惊奇”的情感,是一个情感词汇。倾向性词汇只是情感词汇的一部分,由于情感词的范围太广,基准词的选取存在一定的困难。而且情感词汇相对于褒贬二义倾向性词汇而言更复杂,除了 PMI 还需要其他的语言学特征,才能识别得比较准确,而搜索引擎只能返回点击数和少量的文摘片段,不利于语言学特征的提取。

就 PMI 而言,Manning在《Foundations of Statistical Natural Language Processing》中提到<sup>[16]</sup>:“互信息是衡量独立性的一种很好的方法,但是不是衡量依赖性的一种很好的方法。因为对于依赖性来说,互信息的值由单独词的频率决定。在其他条件相等的情况下,由低频率词组成的二元组的互信息要大于高频率词组成的二元组。”

由于情感词汇的特殊性,不能使用以上的各种方法进行情感词汇的自动获取。本文使用条件随机域(conditional random field, CRF)的机器学习方法,将情感词汇的语法规律作为特征模版,由 CRF 自动生成特征,进行情感词汇的自动获取。实验突破了形容词词性和模式的限制,实现了情感词汇的自动获取。

### 3 情感词汇的一般语法规律

在进行情感词汇获取之前,首先要明确情感词汇的一般语法规律。目前对于情感规律的研究还比较少,仅有的研究也主要集中在句子级和篇章级。句

子级和篇章级的情感规律又和词汇级的情感规律有所不同。例如,在句子级情感计算中,可以考虑句子间连接词的规律,句类的规律和句子的长度规律等等。而词汇的情感计算,由于处理的对象粒度相对较小,因而所能利用的规律也相对较少。在观察大量情感词汇以及情感词汇的上下文后,总结出了以下的情感词汇的一般语法规律。

#### 3.1 词性规律

中文自然语言处理一般都需要进行分词和词性标注的预处理。词汇和词性为后继的各种分析理解提供了基础。其中,词性可以作为情感词汇判断的一个显著特征。

本文使用哈尔滨工业大学的分词和词性标注系统。该标注系统包含的词性为28种。在各个词性中,情感词的分布却不是均匀的,一般形容词、副词、叹词、名词、动词含有情感的比较多,而连词、数词、介词、量词、代词则含有情感的相对较少。由于情感词汇的词性分布不均匀性,可以在情感词汇的识别过程中,使用词性的规律。

#### 3.2 词汇规律

在情感词汇中观察到这样的情况,某些词汇,例如:“的”、“地”、“得”等,常常和情感词汇共同出现,可以作为情感词汇识别的词汇规律。例如:

“隔壁桌上一位热心的华人热情地帮助了他们。”(《少年文摘》2006年第3期,任帝星的意大利音乐之旅)。

“对于零用钱少得可怜的我来说,要赔一个随身听是一件艰巨的事情。”(《少年文艺》2006年第4期,报复)。

“母亲的聪明使父亲微笑得摇头晃脑,全家兴冲冲地等待着四弟露面。”(《文学少年》(初中版)2006年第3期,四弟的绿庄园)。

#### 3.3 词汇重叠规律

词汇的重叠主要集中在动词、形容词、副词。

对于动词的重叠,代表性的观点首先是朱德熙1980年提出的“时量动量”说。他认为“动词重叠式表示动作的量”,即时量和动量。从这个观点他引申出:1)表短时量的重叠式动词用在祈使句里,可以使口气显得和缓些;2)表动量小的重叠式常常表示尝试。其次是刘月华于1983年在朱德熙的基础上提出了“表达功能”说,主要观点是:1)表已然的动作重叠,具有描写作用;2)表未然的动作重叠,具有缓和语气的作用,是委婉表达主观愿望的一种方式;3)有些动词重叠,表示经常性的、反复进行的动作,包含轻松、随便的意味<sup>[17]</sup>。朱德熙和刘月华关于动

词重叠的研究,都说明了动词重叠对于情感表达的影响,无论是语气上隐含的情感,还是主观愿望上直接的情感。

对于形容词的重叠和情感表达的联系,朱德熙在 1956 年曾举了这样的例子:“白白的”有主观的感情色彩,而“白的”没有<sup>[18]</sup>。

对于副词的重叠,一般表达强调修饰,如“匆匆、老老实实、慌慌张张、忙忙碌碌”等,而强调一般都含有某种情感。

### 3.4 否定词与词性搭配规律

否定词与形容词搭配,其中的形容词一般是情感词汇,例如:

“为什么这么不用心?”《少年文摘》,2006 年第 1 期,当时年少春衫薄)。

“爱因斯坦很快与弗莱克斯纳之间发生了一些不愉快。”(《少年文摘》2006 然后第 3 期,令人留恋的学者“大酒店”)。

本文中否定词的获取是通过知网实现的。在知网中选取具有否定意义的义原,如: {neg|否}, {impossible|不会}等,从中抽取出包含否定义原的概念,经人工过滤得到 18 个否定词。

### 3.5 程度副词修饰规律

汉语言中,关于程度副词的分类、各类的作用以及程度副词与其他词性的搭配有很多的研究。其中,陈群在《近代汉语:程度副词研究》<sup>[19]</sup>中指出,汉语的程度副词主要用来修饰性质形容词、某些性状动词以及一部分心理动词。其中,部分性质形容词、部分性状动词和全部的心理动词,都可能是情感词汇。本文的程度副词来自知网,含有 {degree|程度} 义原的副词,经过人工过滤得到 72 个程度副词。

## 4 情感词汇自动获取

将情感词汇的一般语法规律与机器学习方法相结合,实现情感词汇的自动获取,并且根据实验结果,分析以上 5 种语法规律对于情感词汇自动获取的作用,从而找出最佳的特征集用于情感词汇的自动获取。使用的 CRF++ 工具包为情感词汇的自动获取提供了便捷的机器学习工具,可以将更多的精力用于情感词汇本身规律的分析和研究。

### 4.1 条件随机域

条件随机域 (CRF),是一种基于无向图的概率模型,常用于处理序列标注问题,如分词、词性标注等。给定序列  $X$ , 标注序列为  $Y$  的概率公式如式 (1)<sup>[20]</sup>。这里  $x$  是观察值序列,  $y$  是标记序列,  $y/s$  表示子图  $s$  中和  $y$  相连的顶点,  $f_k$  和  $g_k$  都是预先设定的特征。

$$P(y/x) = \exp \left( \sum_{e \in E, k} f_k(e, y/e, x) + \sum_{v \in V, k} u_k g_k(v, y/v, x) \right). \quad (1)$$

情感词汇标注正是一个序列标注问题,当前词是不是情感词不仅仅由词本身决定,而且受当前词上下文的影响。采用条件随机域 CRF++ 工具包,进行情感词汇的自动标注。CRF++ 工具包使用方便,只需要按照工具包要求的格式,整理训练语料和测试语料以及特征模版,CRF++ 就能自动地按照特征模版和训练语料选取特征,并生成相应的模型。

### 4.2 实验步骤

实验语料选自《少年文艺》、《青年文摘》、《新青年》等 9 本文学期刊,选择 2006 年的有强烈感情色彩的文章 515 篇,22 955 个句子,人工地标注其中的情感词汇<sup>[21]</sup>。实验采用 4 倍交叉验证 (4-fold cross validation) 的方法,将数据集分成 4 份,轮流将其中 3 份做训练,1 份做测试,4 次的结果的均值作为对算法精度的估计。交叉验证的方法能够排除单一验证的随机性对实验结果的影响,从而比较客观地体现算法的效果。

以下是情感词汇自动获取的实验步骤:

1) 将上文获得的否定词、程度副词以及人工标注的情感词汇,加入哈工大的分词和词性标注系统的扩展词典,进行分词和词性标注。

2) 为了考察各个特征对情感词汇识别的作用,依次加入情感词汇的语法特征,如果加入的特征提高了实验结果,则保留该特征,反之,则剔除该特征,从而最终寻找出最佳的特征集。实验依次使用了以下 4 种特征集:

特征集 1: 词汇规律和词性规律;

特征集 2: 在特征集 1 的基础上加入重叠规律;

特征集 3: 在特征集 2 的基础上加入否定词与词性搭配规律;

特征集 4: 在特征集 2 的基础上加入程度副词修饰规律。

最终得到特征集 4 的特征模板如表 2 所示,其余特征集的模板根据所包含的情感词汇规律,为特征集 4 的特征模板的相应子集。

3) 根据不同的特征集,按照 CRF++ 要求的格式整理训练语料、测试语料和特征模版。

4) 使用 CRF++ 和以上 4 种特征集,自动标注测试语料中的情感词汇。

5) 分别计算 4 种特征集情况下的正确率和召回率。

6) 实验结果分析。

表 2 情感词汇特征模板  
Table 2 The feature templates of affective words

情感词汇的规律	特征模板项	说 明
词汇	% x[ - 1, 0 ]	情感词的前一个词
	% x[ 1, 0 ]	情感词的后一个词
词性	% x[ - 2, 1 ]	情感词的前面第 2 个词的词性
	% x[ - 1, 1 ]	情感词的前面第 1 个词的词性
	% x[ 0, 1 ]	情感词本身的词性
	% x[ 1, 1 ]	情感词的后面第 1 个词的词性
	% x[ 2, 1 ]	情感词的后面面第 2 个词的词性
	% x[ - 2, 1 ] / % x[ - 1, 1 ] / % x[ 0, 1 ]	情感词与前面第 2 个词,前面第 1 个词的词性的搭配
	% x[ 0, 1 ] / % x[ 1, 1 ] / % x[ 2, 1 ]	情感词与后面第 1 个词,后面第 2 个词的词性的搭配
词汇和词性的搭配	% x[ - 1, 1 ] / % x[ 0, 1 ] / % x[ 1, 1 ]	情感词与前、后词的词性搭配
	% x[ - 1, 0 ] / % x[ 0, 1 ]	情感词的词性与前一个词的搭配
	% x[ 1, 0 ] / % x[ 0, 1 ]	情感词的词性与后一个词的搭配
重叠	% x[ - 1, 2 ]	情感词前一个词的重叠
	% x[ 0, 2 ]	情感词的重叠
否定和词性的搭配	% x[ - 2, 3 ] / % x[ 0, 1 ]	情感词的词性和前面第 2 个词否定的搭配
	% x[ - 1, 3 ] / % x[ 0, 1 ]	情感词的词性和前面第 1 个词否定的搭配
程度副词	% x[ - 2, 4 ]	情感词前面第 2 个词是否为程度副词
	% x[ - 1, 4 ]	情感词前面第 1 个词是否为程度副词

4. 3 实验结果

情感词汇自动获取的结果评测,采用自然语言处理中常用的评测标准:正确率  $P$ ,召回率  $R$  和  $FB1$  值.

$$P = \frac{\text{CorrectEmotionSum}}{\text{EmotionSum}}$$

其中:CorrectEmotionSum 指情感词汇判断正确的数量;EmotionSum 指所有判断为情感词汇的词汇数量.

$$R = \frac{\text{CorrectEmotionSum}}{\text{EmotionStandardSum}}$$

其中:CorrectEmotionSum 指情感词汇判断正确的数量;EmotionStandardSum 指人工标注中情感词汇的总数.

$$FB1 = \frac{2 \cdot P \cdot R}{P + R}.$$

情感词汇自动获取的结果如表 3 所示.

表 3 情感词汇自动获取的实验结果

Table 3 Experiment results in affective word acquisition

特征集	特征数	正确率	召回率	FB1 / %
		$P$ / %	$R$ / %	
特征集 1	346 644	60. 69	54. 93	57. 67

特征集 2	346 654	61. 54	55. 06	58. 12
特征集 3	347 008	61. 22	55. 16	58. 03
特征集 4	346 678	66. 53	54. 60	59. 98

4. 4 实验结果分析

由表 3 可以看出:相比于特征集 1,特征集 2 的结果正确率、召回率、 $FB1$  都有所提高,说明词汇的重叠规律能有效地提高情感词识别效果. 相比于特征集 2,特征集 3 的结果正确率有所下降,但是召回率上升,总体上的  $FB1$  值仍略有下降. 说明否定词和词性的搭配规律不利于定位情感词汇,但有利于找回情感词汇. 特征集 4 的正确率有较大的提高 (5. 31%),但是召回率略微降低 (0. 44%),这说明特征集 4 中包含的程度副词规律,能较准确地定位情感词汇,但是由于程度副词对情感词的限制性,使得召回率下降. 总体上的  $FB1$  值仍然是 4 组特征集中最好的.

对实验的结果进行了进一步地错误分析,找出了导致错误发生的主要原因,主要有以下几个方面:

1)就实验本身而言,实验的每一个步骤都可能造成错误的积累,例如分词和词性标注部分,并不能做到完全的正确. 在否定词和程度副词的集合中,也

并不能确定完全包含所有否定词和程度副词.而且,实验所用的语料是人工标注的,由于精力有限,使得的情感语料规模有限,这也使得统计不够全面.

2)就语言本身而言,汉语语言具备无限丰富性和动态开放性.这使得一方面,本文第 4 部分的情感词汇的规律不能做到全面,也就是不能以有限的规律覆盖无限的语言现象;另一方面,这些规律也不能做到完全的正确,某些符合以上规律的词汇,有时候并不一定是情感词汇.例如:

“我厌恶地看看那只鸡,皱起了眉,朝她挥挥手说:张奶奶,您回家再教训他,我们现在要上课.”(《少年文摘》2006 年第 1 期,一地鸡毛的幸福)——句中的重叠词并不是情感词.

“从此,我在心中再不喊他老师,而是叫他‘崔老头’.”(《少年文艺》,2006 年第 4 期,报复).

“爸爸问为什么,巧巧说,我不喜欢报纸.”(《新青年》,2006 年第 1 期,美人志).——前一句的“不”与动词搭配,动词不表达情感,而后一句的“不”与动词搭配,动词却表达情感.

3)最后,汉语言中的一些词汇是否包含情感,是否是情感词汇,只能通过人工才能判断.目前的机器计算无法判断出来,无论是采用基于规则的,还是基于统计的方法.例如:

“我打了个喷嚏”,“我打了个冷战.”(《少年文摘》,2006 年第 3 期,谁是我的亲母)——其中“喷嚏”和“冷战”,前者不是情感词汇,而后者是情感词汇,只有人工才能区别.

实验最后从语料中随机选取文本 126 篇,包含 130491 个词汇,使用特征集 4 自动获取情感词汇 5382 个,人工过滤,去重得到正确的情感词汇 2607 个.虽然一部分的情感词汇没有被召回,但是将人工过滤的词汇范围从 130491 个减少到 5382 个,大大减少了工作量.

## 5 结束语

本文观察总结了情感词汇的一般规律,在此基础上,使用 CRF 模型实现了情感词汇的自动获取.情感词汇的自动获取大大减轻了人工的工作量,并促进了句子级和篇章级的情感计算.但是本实验只是对于情感词汇自动获取的初步的探索,实验中,情感词汇的一般规律还需要进一步的研究和改进.同时,情感计算是基于语法和语义理解的,因此在情感词汇自动获取中,加入语法和语义分析可能也会有所帮助,这也是下一步要做的工作.

## 参考文献:

- [1] PICARD R W. Affective computing[M]. Cambridge, MA: The MIT Press, 1997: 3-6.
- [2] LU Hugo, SELKER T, LIEBERMAN H. Visualizing the affective structure of a text document[C]//Conference on Human Factors in Computing Systems Florida, USA, 2003: 740-741.
- [3] ZHANG Li, BARNDEN J A, HENDLEY R J, et al. Experimentation in affect detection in open-ended improvisational text[C]//The Annual Meeting of the Association of Computational Linguistics Sydney, 2006: 47-55.
- [4] 董振东,董强.《知网》[EB/OL]. [2007-01-01]. <http://www.keenage.com>.
- [5] 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.  
XU Linhong, LIN Hongfei, PAN Yu, et al. Constructing the affective lexicon ontology[J]. Journal of the China Society for Scientific and Technical Information, 2008, 27(2): 180-185.
- [6] 董大年.现代汉语分类词典[M].上海:汉语大词典出版社,1998: 105-110.
- [7] 王国璋.汉语褒贬义词语用法词典[M].北京:华语教学出版社,2001: 123-128.
- [8] 程志强.中华成语大词典[M].北京:中国大百科全书出版社,2003: 56-60.
- [9] 许小颖,陶建华.汉语情感系统中情感划分的研究[C]//第一届中国情感计算及智能交互学术会议.北京,2003: 199-205.  
XU Xiaoying, TAO Jianhua. The study on affective word classification in Chinese affective systems[C]//The Proceedings of the First Chinese Conference on Affective Computing and Intelligent Interaction Beijing, 2003: 199-205.
- [10] 刘桐菊,于浩,杨沐昀.基于 TFDf 的专业领域词汇获取的研究[C]//第一届学生计算语言学研讨会.北京,2002: 263-267.  
LU Tonghui, YU Hao, YANG Mujun. The research of term extraction in professional field[C]//The Proceedings of the First National Student Workshop on Computational Linguistics Beijing, 2002: 263-267.
- [11] DA LLE B. Study and implementation of combined techniques for automatic extraction of terminology[C]//The 32th Annual Meeting of the Association for Computational Linguistics New Mexico, USA, 1994: 29-36.
- [12] 张晓鹏.汉语特定领域本体的自动构造研究[D].武汉:华中师范大学,2007.  
ZHANG Xiaopeng. The study on automatic construction of ontology in special areas[D]. Wuhan: Huazhong Normal University, 2007.
- [13] HATZIVASSILOPOULOS V, MCKEOWN K R. Predicting

the semantic orientation of adjectives[C]//Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics and the 8th Conference of the European Chapter of the ACL. Madrid, Spain, 1997: 174-181.

- [14] YUEN R W M, CHAN T Y W. Morpheme-based derivation of bipolar semantic orientation of Chinese words[C]//Proceedings of the 20th International Conference on Computational Linguistics. Geneva, Switzerland, 2004: 1008-1014.

- [15] TURNEY P D, LITMAN M L. Measuring praise and criticism: inference of semantic orientation from association [J]. ACM Transactions on Information System (TOIS), 2003, 21(4): 315-346.

- [16] MANNING C D, SCHUTZE H. 统计自然语言处理基础 [M]. 电子工业出版社, 2005: 111-114.

- [17] 吴 晗. 汉语重叠研究综述 [J]. 汉语学习, 2000, 3: 28-33.

- [18] 朱德熙. 现代汉语形容词研究 [J]. 语言研究, 1956, 1: 1-9.

- [19] 陈 群. 近代汉语:程度副词研究 [M]. 四川:巴蜀书社, 2006: 34-41.

- [20] LAFFERTY J, MCCALLUM A, PEREIRA F. Conditional random fields: probabilistic models for segmenting and labeling sequence data[C]//18th International Conf on Machine Learning. San Francisco, USA: Morgan Kaufmann, 2001: 282-289.

- [21] 徐琳宏, 林鸿飞, 赵 晶. 情感语料库的构建和分析 [J]. 中文信息学报, 2008, 22(1): 116-122.

作者简介:



陈建美,女,1985年生,硕士研究生,主要研究方向为情感词汇自动获取和情感词汇消歧.



林鸿飞,男,1962年生,教授,博士生导师,现任《中文信息学报》编委,中文信息学会理事,中国中文信息学会信息检索专业委员会委员,中国人工智能学会离散数学专业委员会副主任,中国人工智能学会机器学习专业委员会委员.主要研究方向为搜索引擎、文本挖掘、情感计算和自然语言处理.主持多项国家自然科学基金和863计划项目,发表学术论文100余篇.



杨志豪,男,1973年生,副教授,博士研究生,主要研究方向为文本挖掘和中文信息处理,发表学术论文20余篇.

## 第2届 Pao-Lu Hsu会议:机器学习与计算识别 The Second Pao-Lu Hsu Conference (2009): Machine Learning and Computational Recognition

The Second Pao-Lu Hsu Conference (2009): Machine Learning and Computational Recognition will be held on Xi'an Jiaotong University in Xi'an, China in the last week of June, 2009 (June 22 ~ 28). This is a joint event of the Institute for Information and System Science of Xi'an Jiaotong University (Zongben Xu Lab) and the Microsoft Lab on Statistics and Information Technology at Peking University (PKU MSRA Lab). This conference intends to promote machine learning and Computational Recognition in China. The audiences are expected to be faculty and students in Machine Learning and related fields from both inside and outside China.

### ORGANIZING COMMITTEE

Bin Yu, Chair of organizing committee, UC Berkeley

Zongben Xu, co-Chair, Xi'an Jiaotong University

Hang Li, Microsoft Research Asia

Ming Jiang, Peking University

Zhi Geng, Peking University

Jiangshe Zhang, Xi'an Jiaotong University

Web site: <http://icnp.xjtu.edu.cn/meeting.php?blogId=6>