

DOI:10.3969/j.issn.1673-4785.201301029

网络出版地址: <http://www.cnki.net/kcms/detail/23.1538.TP.20130603.1602.009.html>

## 情感倾向判断中基准词的选择

程传鹏,王海龙

(中原工学院 计算机学院,河南 郑州 450007)

**摘要:**针对已有研究工作中基准词选择的不足之处,提出了一种情感倾向判断中基准词选择的方法.依照基准词的定义,分别从情感度、情感倾向度、情感歧义性 3 个方面考虑,选择出数量尽可能少、最具有代表性的情感词作为基准词.首先从《知网》所发布的情感词语中,筛选出最初的候选基准词,计算出这些候选基准词的情感度.然后分别计算出情感度排名靠前的正面情感词和负面情感词的情感倾向度.最后选择情感倾向度较大的词语,作为最终的基准词.实验结果表明,按照文中所选择的基准词所得到的情感倾向判断的准确率较高.

**关键词:**基准词;情感度;情感倾向度;情感歧义性

**中图分类号:** TP391 **文献标志码:** A **文章编号:** 1673-4785(2013)04-0349-07

中文引用格式:程传鹏,王海龙. 情感倾向判断中基准词的选择[J]. 智能系统学报, 2013, 8(4): 349-355.

英文引用格式:CHENG Chuanpeng, WANG Hailong. Research on selection of paradigm words in the judgment of emotional tendency[J]. CAAI Transactions on Intelligent Systems, 2013, 8(4): 349-355.

## Research on selection of paradigm words in the judgment of emotional tendency

CHENG Chuanpeng, WANG Hailong

(School of Computer Science, Zhongyuan Institute of Technology, Zhengzhou 450007, China)

**Abstract:** In light of the weakness of standard words selection, which exists in previous research studies, a method for selecting standard words in the judgment of emotional tendency was proposed in this paper. By considering three key aspects, which include degree of emotion, tendency of emotion and ambiguity of emotion, the most representative typical sentiment words were chosen to act as standard words based on the definition of standard words. Firstly, initial standard words were screened from emotion words issued in HowNet, the degree of emotion of these candidate standard words was computed, then the emotional tendency degree of the positive emotion words and negative emotion words, which have high ranking of emotion degree, was calculated, respectively. Finally, the larger emotional tendency words were used as final standard words. The results of test shows the methods used in this paper can gain a high accuracy for judging the tendency of emotion.

**Keywords:** standard words; degree of emotion; tendency of emotion; ambiguity of emotion

基准词集是指褒贬义倾向非常明显、强烈、具有代表性的词汇所构成的集合<sup>[1]</sup>.情感倾向判断中,基准词是衡量其他词语情感倾向的一个参照物,因此,基准词的选择对正确地判断情感的倾向有着至关重要的作用.在文献[2]中,Turney 通过分析词汇上下文信息研究其情感倾向,采用 PMI-IR 方法,使用 2 个词汇作为种子来判断其他短语的语义倾向.之后又将单对种子扩展成多对种子,选取了正反面各 7

个词汇作为情感基准词.文献[3]首先从《知网》中选择出 2 146 个褒义词和 3 299 个贬义词,从这些词语中选择没有褒贬歧义的常用词语作为测试集,然后将测试集中的词语在 Google 中搜索,返回 hits 数,即它们在 Web 上的词频降序排列,选取词频最高的 280 个词语作为基准词.文献[4]将文献[3]中所选出的基准词中语义重复的词替换成新的、有较高 hits 数的褒义词或贬义词,最后得到新的 40 组褒贬基准词,这 40 组褒贬基准词的特点是在保留较高的使用频率外,排除了语义相同的情况,提高了词汇的覆盖面.文献[5]通过词的聚类,将中文词聚类应

用到基准词的选择中,从构建好的目标领域本体中选出一组初始的种子词,通过词的扩展聚类得出下一代种子词,然后通过迭代得出优化的基准词.文献[6]利用公式计算语料库中名词、形容词和动词的类别区分能力,选出区分能力较强的词  $M$  个,用所选出的词与已有的情感词表做交集,选择语料中出现频率高的前  $N$  个词作为最终的基准词.文献[6]使用的基础情感词以《知网》发布的情感词语集为基础,通过人工挑选,去掉一些不太常用或者情感倾向不很明显的词语,最后得到褒义词3 219个,贬义词2 905个,最终的基础情感词词典包含6 196个基础情感词.然后计算它们的情感倾向权值,去掉分类不正确的词以及权值过低的中性词,最后得到5 281个基础情感词.文献[7]选择基准词的方法是:从《知网》<sup>[8]</sup>所提供的情感词语中选择全部义项只有一种极性的形容词和名词作为基准词,并且认为单一义项的词语不会出现褒贬歧义和情感极性较弱的情况.

通过人工选择具有倾向性的基准词方法的优点是选择的基准词准确率高,但由于人知识的有限性和片面性,选择的基准词往往会漏选、错选.通过已经制定好的褒贬倾向性词典,然后利用同义词词典对词典进行扩展的方法,虽然所选择的基准词较全面,但给后面的倾向性判断带来大量的计算.由于搜索引擎的更新,利用搜索引擎所返回来的 hits 数筛选基准词的方法,实际上也是不够客观的.

## 1 基准词的选择方法

一个词语要作为基准词,一般应该满足3个条件:情感代表性强、情感歧义少、情感极性强.下面分别从这3个因素来叙述本文所提出的基准词选择的方法.

### 1.1 情感语义相似度

通过改进的词语相似度计算方法,来选择情感歧义少的情感词作为基准词.文献[9]基于《知网》的相似度计算,依照《知网》对义原的描述,文献[9]将所有的义原用一个树状的层次结构图来表示,如图1所示.

由于所有的义原根据上下位关系构成了一个树状的义原层次体系,假设2个义原在这个层次体系中的路径距离为  $d$ ,文献[9]采用式(1)来计算2个义原之间的语义距离:

$$\text{sim}(p_1, p_2) = \frac{a}{d + a}. \quad (1)$$

式中各参数的含义见文献[9].式(1)的计算,并没

有考虑义原在语义树状图中的深度.依照树状图的定义,越是处于层次低的义原节点,概念描述的越具体,比如“兽”和“走兽”,而处在层次越高的义原节点概念描述越笼统,比如“万物”和“物质”.因为对于  $d$  值相同的2对义原,处于语义层次树较低层次的应该比处于较高层次的2对词语的相似度要大.因此,把式(1)改进为

$$\text{sim}(p_1, p_2) = \frac{a + H(p_1) + H(p_2)}{a + H(p_1) + H(p_2) + L(p_1, p_2)}. \quad (2)$$

式中:  $H(p_i)$  为义原节点到根节点的路径长度,  $L(p_i, p_j)$  为2个义原节点第1次到达同一个父节点所经过的最长路径长度.



图1 树状的义原层次结构

Fig.1 Arborecence hierarchical structure of semantic unit

假设情感词  $W_1$  的义项为  $\{S_{11}, S_{12}, \dots, S_{1n}\}$ , 情感词  $W_2$  的义项为  $\{S_{21}, S_{22}, \dots, S_{2m}\}$ , 义项之间的相似度计算, 依然参照文献[9]中的计算方法, 词语  $W_1$  和  $W_2$  的相似度为各个义项的相似度之最大值, 如式(3)所示:

$$\text{sim}(W_1, W_2) = \max_{i=1,2,\dots,n, j=1,2,\dots,m} \text{sim}(S_{1i}, S_{2j}). \quad (3)$$

如果直接将式(3)用于情感倾向判断中的词语相似度计算, 实际上是不太合理的. 具有多个义项的词语可能会在情感倾向上产生歧义, 也就是说词语的某个义项可能为正面情感倾向, 而另外一个义项可能为负面情感倾向. 在选择基准词的时候, 应该尽量排除掉具有情感歧义的词. 因此本文在计算情感词语语义相似度时采用情感词  $W_1$  和  $W_2$  的各个义项的语义相似度之和的平均值, 假设情感词  $W_1$  有  $n$  个义项,  $W_2$  有  $m$  个义项, 如果用  $E(W_1, W_2)$  表示  $W_1, W_2$  情感词语语义相似度, 其计算公式可以表示成式(4):

$$E(W_1, W_2) = \frac{\sum_{i=1,2,\dots,n, j=1,2,\dots,m} \text{sim}(S_{1i}, S_{2j})}{n \times m}. \quad (4)$$

### 1.2 情感度的计算

情感度指情感词能够代表负面情感词或者正面情感词的程度. 通过对情感词的情感度的计算, 选择出一些最具有代表性的情感词作为基准词. 基准词

的情感倾向应该特别明显,而且从计算量上考虑,情感词越少越好.应该从正面情感词和负面情感词中选择出最具有倾向性代表意义的词语来作为基准词.

假设在情感词表中有另外的5个情感词  $S_1, S_2, S_3, S_4, S_5$ , 它们与  $W_1, W_2$  的情感语义相似度如表1所示.表1中,  $\text{sim}(S_i, W_j) \neq 1$ , 因为如果相似度等于1的话,那么可能2个词为同义词或者近义词.从表1中可以看出,  $W_1$  与其他5个情感词的情感语义相似度都要大于  $W_2$  与其他5个情感词的情感语义相似度,直观地判断,应该选择  $W_1$  为基准词.

表1 基于知网的词语相似度计算

Table 1 Similarity calculation of words based on HowNet

词语	$W_1$	$W_2$
$S_1$	0.85	0.82
$S_2$	0.92	0.89
$S_3$	0.86	0.84
$S_4$	0.81	0.76
$S_5$	0.86	0.83

对于所有的情感词表,分别计算两两之间的情感语义相似度,如果用  $S(W_i, W_j)$  表示  $W_i, W_j$  情感语义相似度,那么所有词的情感语义相似度可以表示成如表2所示.

表2 情感词之间的语义相似度

Table 2 Semantic similarity between emotion words

词语	$W_1$	$W_2$	...	$W_i$	...	$W_n$
$W_1$	1	$S(W_1, W_2)$	...	$S(W_1, W_i)$	...	$S(W_1, W_n)$
$W_2$	$S(W_1, W_1)$	1	...	$S(W_2, W_i)$	...	$S(W_2, W_n)$
$\vdots$	...	...	...	...	...	...
$W_i$	$S(W_i, W_1)$	$S(W_i, W_2)$	...	1	...	$S(W_i, W_n)$
$\vdots$	...	...	...	...	...	...
$W_n$	$S(W_n, W_1)$	$S(W_n, W_2)$	...	$S(W_n, W_i)$	...	1

如果  $S(W_i, W_j) \geq \alpha$ , 建立  $W_i$  和  $W_j$  之间的关系,处理完所有的情感词后,形成如下图2所示.

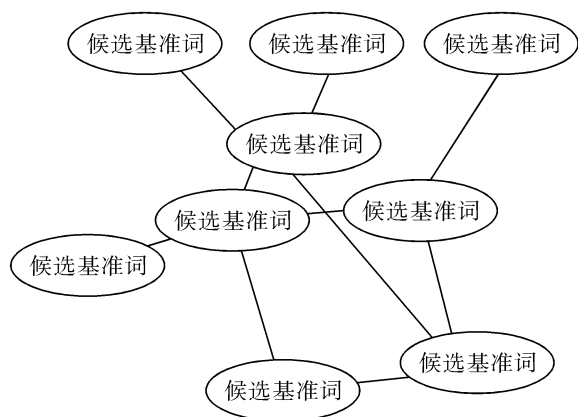


图2 基准词语义连接

Fig.2 Semantic connection diagram of standard words

**定义1** 情感度(emotion degree)指的是候选基准词能够代表正面情感词或者负面情感词情感的程度,用  $D_1(W_i)$  表示.

**定义2** 词对连接强度  $C(W_1, W_2)$  指的是2个词语连接的紧密程度,用2个词语的情感语义相似度表示,即  $C(W_1, W_2) = E(W_1, W_2)$ .

**定义3**  $W_i$  的邻接词条指的是指满足  $C(W_i,$

$W_j) \geq \lambda$  并且  $i \neq j$  的所有  $W_j$  的集合.

基准词指的是情感表现力强,最能够体现正面情感或者负面情感的词条,依据此定义,一个情感词要作为基准词,应该存在以下2个事实:

1) 如果一个词条与情感词集合中其他词条都有较高的连接度,那么该词条的情感度高;

2) 如果一个词条的情感度高,那么与该词条连接度高的词条的相应的情感度也高;

设  $W_i$  是某个候选基准词,其情感度为  $D_1(W_i)$ , 记  $F_j$  是所有  $W_i$  的所有邻接词条的集合,  $n_j = |F_j|$  是指满足  $E(W_k, W_j) \geq \alpha$  的所有  $W_k$  的总数,依据上面2个事实,候选基准词的情感度可以用式(5)进行计算:

$$D_1(W_i) = \sum_{W_j \in F_i} \frac{D_1(W_j)}{n_j}. \quad (5)$$

定义词语语义的邻接矩阵:

$$G = \begin{bmatrix} g_{11} & \cdots & g_{1n} \\ \vdots & & \vdots \\ g_{n1} & \cdots & g_{nn} \end{bmatrix}. \quad (6)$$

式中:

$$g_{ij} = \begin{cases} 1, & \text{if } L(W_i, W_j) \geq \lambda; \\ 0, & \text{if } i = j; \\ 0, & \text{if } L(W_i, W_j) < \lambda. \end{cases}$$

在邻接矩阵  $G=(g_{ij})$  定义的基础上,式(6)可写成:

$$D_1(W_i) = \sum_{W_j \in F} \frac{g_{ij}}{n_j} D_1(W_j). \quad (7)$$

如果情感词表中有  $n$  个候选基准词,它们的情感度可以用一个  $n$  维的列向量  $P$  来表示,即

$$P = [D_1(W_1) \quad D_1(W_2) \quad \cdots \quad D_1(W_n)]^T.$$

则式(7)可写成:

$$P = A^T P. \quad (8)$$

式中:  $A_{ij} = g_{ij}/n_j$ .

式(8)是递归定义的,因此可以用迭代的办法求解.可以看出,候选基准词的情感度  $P=(D_1(W_i))$  实为矩阵  $A^T$  对应于特征根为 1 的特征向量,当  $A^T = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$  时,将得到  $P_1=P_2$ ,无法比较候选基准词的情感度的大小.因此,要对式(8)进行改进.采用加权算法如式(9):

$$P = \frac{(1-d)E}{n} + dA^T P. \quad (8)$$

式中:  $n$  为候选基准词个数;  $E$  为  $n \times n$  的矩阵,矩阵的元素值都是 1;  $d \in [0, 1]$ , 参照文献[10],文中  $d$  的取值为 0.85.

### 1.3 情感倾向度的计算

通过对情感词情感倾向度的计算,选择情感倾向较大的词来作为基准词.设  $W_{n_1}, W_{n_2}, \dots, W_{n_{1000}}$  为上一步所选择的负面基准词,个数为 1 000.  $W_{p_1}, W_{p_2}, \dots, W_{p_{1000}}$  为 1 000 个正面情感基准词.如果直接用刘群所开发的相似度计算软件,会发现一些情感倾向相反的词语,语义相似度的值却很大,比如一对明显语义相反的词语“小气”和“大方”,按照刘群的计算方法  $\text{sim}(\text{“小气”}, \text{“大方”}) = 0.81$ ,通过查看义项发现“大方”的义项是“ADJ avalue | 属性值, tolerance | 气量, generous | 慷慨”,“小气”的义项是“ADJ avalue | 属性值, tolerance | 气量, miser | 吝啬”刘群所开发的软件是基于《知网》的,《知网》是一个以汉语和英语的词语所代表的概念为描述对象,以揭示概念与概念之间以及概念所具有的属性之间的关系为基本内容的常识知识库<sup>[9]</sup>.在《知网》中,义原之间除了上下位关系外,还有很多种其他的关系,如果在计算时考虑进来,可能会得到更精细的义原相似度度量,义原一方面作为描述概念的最基本单位,另一方面,义原之间又存在复杂的关系.《知网》认为,具有反义或者对义关系的 2 个义原比较相似,因为它们在实际的语料中可以互相替换的可能性很大.因此在判断情感倾向时,如果 1 个负面情感词和 1 个

正面情感词语义相似度高的话,说明它们二者之间的情感倾向距离大.如果用  $D_2(W)$  表示词语  $W$  的情感倾向度,那么正面情感词  $W_p$  的情感倾向度可以用该词与所有的负面情感词的语义相似的和来表示,负面情感词  $W_n$  的情感倾向度可以用该词与所有的正面情感词的语义相似的和来表示,计算方法为

$$\begin{cases} D_2(W_p) = \sum_{i=1}^n E(W_p, W_{n_i}), \\ D_2(W_n) = \sum_{i=1}^n E(W_n, W_{p_i}). \end{cases}$$

## 2 基准词选择过程以及结果

1) 从 2010 年《知网》所发布的 3 730 个正面评价词语和 3 116 个负面评价词语中,去掉单字和不常见的词,结果得到 3 256 个正面情感词和 2 986 个负面情感词,以这些情感词作为最初的候选基准词;

2) 由前面方法分别计算正面情感词和负面情感词的情感度;

3) 分别从正面情感词和负面情感词中选择出情感度排名靠前的 1 000 个词;

4) 由前面方法分别计算出 1 000 对正面情感词和负面情感词的情感倾向度;

5) 选择情感倾向度较大的词语,作为最终的基准词;

经过前面 5 步的选择后,排名靠前的 40 对正面情感词和负面情感词分别如下所示:

“良好、美好、最佳、上等、容易、最好、美丽、顶级、宽大、精彩、快乐、端正、稳定、优秀、高级、确切、明亮、热情、清新、出色、大方、便宜、积极、著名、灵活、牢固、真实、简单、超级、必需、先进、纯朴、轻快、欢乐、仁爱、平坦、聪明、出色、平安、成熟”.

“暗淡、昂贵、傲慢、薄弱、悲观、病弱、不当、愚蠢、脏乱、自大、糟糕、罪恶、杂乱、愚昧、有害、阴冷、虚假、严重、阴暗、凄凉、消极、下等、无效、无情、强制、凶狠、颓废、贪心、缺德、奢侈、散乱、危险、穷困、破旧、萧条、小气、轻狂、失常、俗气”.

为了进一步丰富基准词,本文通过哈尔滨工业大学信息检索研究室所提供的《同义词词林扩展版》对所选择的负面情感基准词和正面情感基准词进行同义词扩充.结构如图 3 所示.

为了避免出现某些词和基准词语义相似度高而词义却相反所带来的误判,对未知词语的情感判断时,首先查找基准词以及基准词的同义词,如果没有找到,则计算未知词语和基准词的语义相似度.

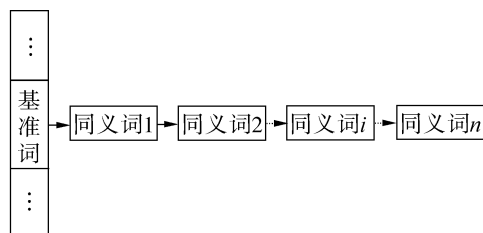


图3 基准词同义扩展

Fig.3 Synonym expansions of standard words

### 3 实验及评价

从网上的评论中搜集到4个领域的中文数据集,分别是:国内某知名网上商城中对某款电子产品的用户评价作为测试语料1,新浪网财经论坛中对某只股票的用户评价作为测试语料2,携程旅行网上对某个酒店的用户评价作为测试语料3,搜狐体育上对某个比赛规则的用户评价作为测试语料4。其中,测试语料1包含正面评价3 690条,负面评价4 721条;测试语料2包含正面评价4 215条,负面评价3 976条;测试语料3包含正面评价2 876条,负面评价3 128条;测试语料4包含正面评价4 134条,负面评价3 157条。以本文所选择的基准词与文献[2]所选择的基准词和文献[3]所选择的基准词进行情感判断比较。其中文献[2]的基准词是由人工选择的,所选择的40对基准词如下所示:

“健康、安全、天下第一、美丽、超级、保险、卫生、天使、英雄、精选、快乐、权威、稳定、优秀、高级、精英、最好、最佳、幸福、容易、高手、文明、积极、著名、漂亮、完美、真实、简单、和平、开通、先进、便宜、优质、欢乐、美好、良好、不错、出色、成熟、完善”。

“合作、黑客、疯狂、错误、事故、非法、失败、背后、麻烦、不良、病人、恶意、色情、暴力、黄色、浪费、落后、漏洞、有害、讨厌、自负、不安、魔鬼、花样、野蛮、陷阱、不当、腐败、无情、失误、淫秽、流氓、虚假、残酷、变态、脆弱、不合格、愚人、恶劣、恶魔”。

文献[3]中所选择的基准词是通过词典进行构造的,如下所示:

“好、安全、不错、喜欢、加速、舒适、豪华、满意、爱、解决、风格、优势、保证、全新、实在、舒服、稳定、方便、品质、提升、乐趣、省油、先进、成功、漂亮、最好、保护、好车、值得、良好、满足、享受、出色、提高、适合、平稳、轻松、优点、完美、实用”。

“碰撞、噪音、事故、毛病、不好、严重、下降、缺点、不够、死、不足、故障、缺陷、郁闷、撞击、断裂、失望、担心、倒、车祸、遗憾、怀疑、不行、变形、断、危险、震动、损失、噪声、麻烦、冲击、隐患、后悔、恐怕、粗糙、颠簸、造成、难看、不爽、伤害”。

采用KNN分类的方法,以这些情感基准词作为KNN分类器的特征词,特征词的加权方法为

$$W(t, d) = \frac{tf(t, d) \times \log(N/n_t)}{\sqrt{\sum_{t \in d} [tf(t, d) \times \log(N/n_t)]^2}}$$

式中: $W(t, d)$ 为情感词 $t$ 在在语料 $d$ 中的权重,而 $tf(t, d)$ 为 $t$ 在 $d$ 中的词频, $N$ 为文档的总数, $n_t$ 为语料中出现 $t$ 的文档数。

试验中,词语相似度的计算公式 $\sum_{i=1}^3 \beta_i \text{sim}_i(S_1, S_2)$ 中 $\beta_1=0.6, \beta_2=0.2, \beta_3=0.2, \text{sim}_s(P_1, P_2)$ 中 $\lambda_1, \lambda_2$ 取了3组值:

1)  $\lambda_1=0.6, \lambda_2=0.4$ ;

2)  $\lambda_1=0.5, \lambda_2=0.5$ ;

3)  $\lambda_1=0.4, \lambda_2=0.6$ 。

$\text{sim}_s(P_1, P_2)$ 公式中 $\alpha_1, \alpha_2$ 也分别取了3组值:

1)  $\alpha_1=0.6, \alpha_2=0.4$ ;

2)  $\alpha_1=0.7, \alpha_2=0.3$ ;

3)  $\alpha_1=0.5, \alpha_2=0.5$ 。

在构造邻接矩阵 $G$ 时, $\lambda$ 取值为0.75。

用的评价指标为

$$\text{准确率} = \frac{N_c}{N_T} \times 100\%$$

式中: $N_c$ =能正确找出文本倾向的总数; $N_T$ =测试文本的总数。KNN分类中 $K$ 的取值分别为1 000、1 200、1 400、1 500,分别采用40组人工选择的基准词对、由词典构造的基准词对和本文所选择的40组情感词对分别作为特征词,对4组语料进行情感倾向分类。分类准确率的结果比较如图4~7所示。

由于本文在选择基准词时,考虑到了情感词的情感歧义、情感度、情感倾向性。本文所提出的方法在准确率上都要优于其他2种方法,因而本文的方法具备有一定的实用性。图4~7的实验数据的比较结果,也验证了该结论。

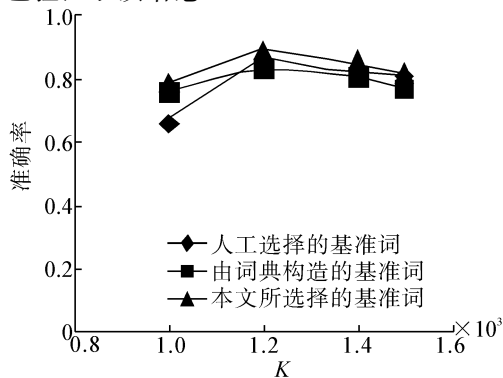


图4 对测试语料1的准确率比较

Fig.4 Accuracy comparison for test corpus 1

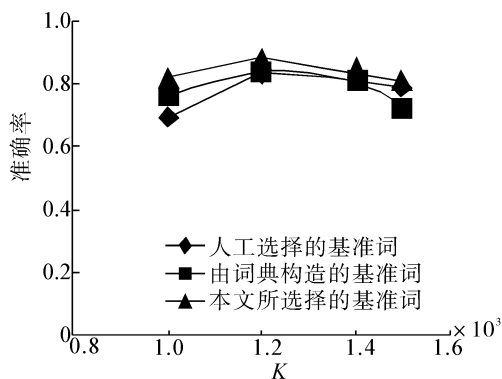


图5 对测试语料2的准确率比较

Fig.5 Accuracy comparison for test corpus 2

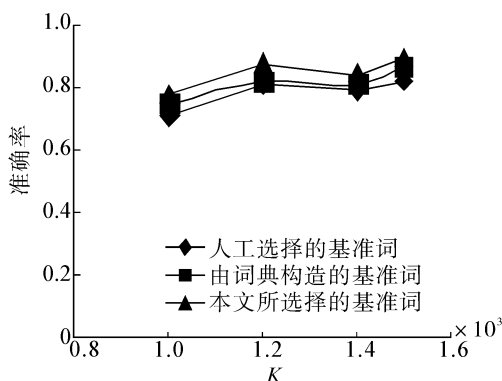


图6 对测试语料3的准确率比较

Fig.6 Accuracy comparison for test corpus 3

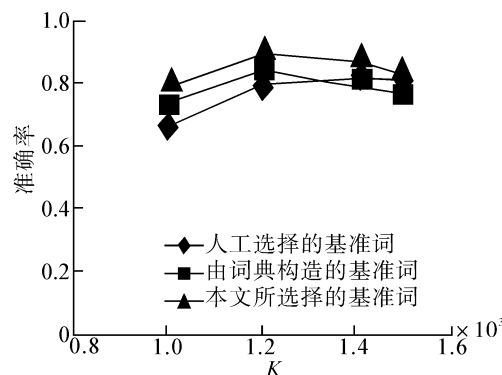


图7 对测试语料4的准确率比较

Fig.7 Accuracy comparison for test corpus 4

为了验证不同的基准词个数对情感分类准确率的影响,分别从本文所得到的基准词中选择出排名靠前的15、20、25、30、40、50、60、75、80、85、90、95、100、120对基准词,对4组测试语料分别测试,取其平均值。得到的准确率如图8所示。从图8可以看出,基准词的数量从15~40时,情感倾向的分类的准确率得到了很快的提高,其中,在选择40对基准词时,准确率为85%;基准词的数量大于40之后,准确率

虽然得到了提高,但总体上来看变化趋势较为缓慢。在实际应用中,随着基准词数量的增多,计算量也会增大。从准确率和时间开销综合考虑,应该选择的基准词数量在40~60。

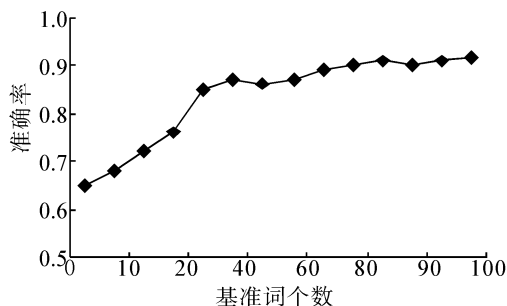


图8 不同基准词个数对准确率的影响

Fig.8 Influence of the number of different standard words on accuracy rate

## 4 结论

本文分析了已有工作中基准词选择的优点和缺点,在此基础上,提出了一种情感倾向判断中基准词选择的方法。分别考虑到了情感词的情感歧义性、情感代表性、情感倾向性,并相应地给出了计算公式。通过实验测试表明,本文所提出的基准词,在情感倾向的判断上准确性较高。文章的主要贡献有:

- 1) 对词语相似度方法进行了改进,通过计算情感词的情感语义相似度,排除掉一些有情感歧义的词作为基准词的可能。
- 2) 通过计算情感度,选择出情感代表性强的词语作为候选基准词。
- 3) 通过对情感词情感度的计算,选择出情感倾向性强的情感词作为候选基准词。

实验结果表明,采用本文选择的基准词,在情感倾向判断的准确率上,要优于传统的方法。由于词典资源的不完备性,以及网络上的新词,比如“二”、“浮云”、“伪娘”等,都没有在词典中出现,以后的研究中,将进一步的考虑这些未登录词的语义对文本相似度的影响。

## 参考文献:

- [1] 孙春华, 刘业政, 彭学仕. 一种含强度的基准词选择和词汇倾向性判别方法[J]. 情报学报, 2011, 30(12): 1261-1267.  
SUN Chunhua, LIU Yezheng, PENG Xueshi. A method for paradigm words selection with intensity information and word sentiment orientation discrimination[J]. Journal of the China Society for Scientific and Technical Information, 2011, 30(12): 1261-1267.
- [2] PETER D T. Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews[C]//

- Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics. Philadelphia, USA, 2002: 417-424.
- [3] 朱嫣岚, 闵锦, 周雅倩, 等. 基于 HowNet 的词汇语义倾向计算[J]. 中文信息学报, 2006, 20(1): 14-20.  
ZHU Yanlan, MIN Jin, ZHOU Yaqian, et al. Semantic orientation computing based on HowNet[J]. Journal of Chinese Information Processing, 2006, 20(1): 14-20.
- [4] 杨昱昂, 吴贤伟. 改进的基于知网词汇语义褒贬倾向性计算[J]. 计算机工程与应用, 2009, 45(21): 91-93.  
YANG Yubing, WU Xianwei. Improved lexical semantic tendentiousness recognition computing[J]. Computer Engineering and Applications, 2009, 45(21): 91-93.
- [5] 彭学仕, 孙春华. 面向倾向性分析的基于词聚类的基准词选择方法[J]. 计算机应用研究, 2011, 28(1): 114-116.  
PENG Xueshi, SUN Chunhua. Paradigm words selecting method based on word clustering for sentiments analysis[J]. Application Research of Computers, 2011, 28(1): 114-116.
- [6] 柳位平, 朱艳辉, 栗春亮, 等. 中文基础情感词词典构建方法研究[J]. 计算机应用, 2009, 29(10): 2875-2877.  
LIU Weiping, ZHU Yanhui, LI Chunliang, et al. Research on building Chinese basic semantic lexicon[J]. Journal of Computer Applications, 2009, 29(10): 2875-2877.
- [7] 宋乐, 何婷婷, 王倩, 等. 极性相似度计算在词汇倾向性识别中的应用[J]. 中文信息学报, 2010, 24(4): 63-67.
- SONG Le, HE Tingting, WANG Qian, et al. Application of polarity similarity in word semantic orientation identification [J]. Journal of Chinese Information Processing, 2010, 24(4): 63-67.
- [8] 董振东, 董强. 知网[DB/OL]. [2012-09-25]. <http://www.keenage.com>.
- [9] 刘群, 李素建. 基于《知网》的词汇语义相似度的计算[C]//第3届汉语词汇语义学研讨会. 台北, 中国, 2002.
- [10] 方芳, 李仁发, 何建军. 基于改进 PageRank 的 BA 演化模型[J]. 计算机工程与设计, 2010, 31(9): 1901-1904.  
FANG Fang, LI Renfa, HE Jianjun. BA evolution model based on improved PageRank[J]. Computer Engineering and Design, 2010, 31(9): 1901-1904.
- 作者简介:**
- 

程传鹏, 男, 1977 生, 讲师, 主要研究方向为自然语言处理、机器学习。



王海龙, 男, 1976 生, 副教授, 主要研究方向为图像处理、机器学习。

## 第 8 届 IEEE 国际智能信号处理研讨会

### 8th IEEE International Symposium on Intelligent Signal Processing

The 8th IEEE International Symposium on Intelligent Signal Processing will take place in the enchanting Funchal, centre of Portugal's Madeira, a venue chosen to enhance interaction between participants, as well as give the event an appropriate setting.

WISP 2013 aims to bring together scientists, researchers, and engineers to present the latest achievements and innovations related to intelligent signal processing in complex, embedded, and real time systems.

#### TOPICS

Papers are solicited on all information processing aspects related to the theory and application of computer based signal processing. Attention will be focused on, but not limited to: analog, digital, and mixed signal processing, image processing, computer vision, pattern recognition, information retrieval, bioengineering, machine learning, man-machine interaction, robotics, modeling, diagnostics, control, uncertainty-handling, artificial intelligence, soft computing techniques, anytime algorithms, emerging technologies, adaptive and nonlinear processing, intelligent signal processing, sensors, embedded systems, real-time systems, distributed measurement, diagnostics, and control systems, system identification and control, information fusion, VLSI/ASIC for intelligent signal processing, tool integration for better performance, performance modeling, performance measurement, virtual systems, applications.

**Website:** <http://trivent.hu/WISP2013/>