

# 基于语境歧义词的句子情感倾向性分析\*

宋艳雪, 张绍武, 林鸿飞

大连理工大学 计算机科学与技术学院, 大连 116024

E-mail: songyanxue86@gmail.com

**摘要:** 目前, 在自然语言处理领域, 关于词语搭配方面的研究主要是在语料基础上抽取搭配, 本文从情感的角度研究语境歧义词的搭配, 这种搭配对文本情感倾向性分析方面具有实际重要的意义。首先使用关联规则挖掘的方法确定语境歧义词候选搭配集, 然后通过 PMI 过滤后判断每对搭配词是否具有情感倾向性, 最终构建语境歧义词搭配词典。我们采用语义分析的方法, 在构建的语境歧义词搭配词典基础上对句子进行情感倾向性分析。通过在 COAE2008 语料集和情感语料库上进行实验, 证明了在判断句子情感倾向性时考虑到语境歧义词的重要性, 其对句子进行情感倾向性判断的正确率有很大的影响。

**关键字:** 语境歧义词; 关联规则; 倾向性; 情感计算

## Sentiment Analysis of Sentence Based on Ambiguous Words

Song Yanxue, Zhang Shaowu, Lin Hongfei

School of Computer Science and Technology, Dalian University of Technology, Dalian 116024

E-mail: songyanxue86@gmail.com

**Abstract:** Currently, in natural language processing, the research on the collocation words is mainly about abstracting collocations from corpus. In this paper, we will study the ambiguous collocations from the aspect of analyzing sentiment, this collocations are important for text analyzing sentiment. First, we will use Association Rule to determine if word pairs are collocations, and cut off some ones that are not necessary though PMI, then determine whether the collocations have ambiguity, after that, we will built a ambiguous collocations dictionary on which we determine sentences' tendency by the method of Semantic Analysis. Experiments we worked on the COAE2008 corpus showed that ambiguous collocations is very important and essential on the sentiment analysis, it can make a large influence on the accuracy of sentiment analysis.

**Keywords:** ambiguous collocations; association rule; orientation; affective computing

### 1 前言

目前, 在分析文本情感倾向性时, 出发点通常是情感词。然而自然语言中的多数词汇本身并不带有明显的感情色彩, 在应用到实际文本中时, 一些词汇会由于具体的语境或不同的搭配关系表达出褒义或者贬义色彩。本文中的“语境歧义词”是指句子中的一些词汇本身没有明显的情感倾向性, 但是当它们与上下文中的某些词语搭配后, 产生了明显的情感倾向性。例如:

句子 1: 这学期女儿学习成绩突出, 被当选为学习班长, 我们决定给她买一个芭比娃娃。

句子 2: 由于长期在电脑旁工作, 老李得了颈椎突出, 现在不能正常的工作和生活了。

“突出”本身并不能表达明显的情感倾向性, 但是当出现“成绩+突出”和“颈椎+突出”搭配后, “突出”表现出两种不同的情感倾向性, 我们把类似“突出”这样的词汇称为语境歧义词。语境歧义词不能单纯地通过添加到情感词典中去分析, 需要凭借其所处语境来判断其倾向性。

当前, 国内外对词语搭配的研究, 主要是从语言学角度出发去挖掘词语搭配效果与认知功能的研究。Igor A. Bolshakov<sup>[1]</sup>通过统计学方法从语料库中抽取正确搭配, 构建一个搭配词典, 为搭配领域做出了贡献, 该词典还没有考虑情感因素。车万翔、刘挺<sup>[2]</sup>通过计算语料中出现的词对的“搭

\* 基金项目: 国家自然科学基金资助项目 (编号: 60673039, 60973068)、国家 863 高科技计划资助项目 (编号: 2006AA01Z151)、教育部留学回国人员科研启动基金和高等学校博士学科点专项科研基金资助课题 (编号: 20090041110002)。

配强度系数”值来衡量这种搭配关系的强弱。目前的研究,很少有从情感的角度出发,去挖掘语境歧义词搭配方面的情感倾向性问题。然而,搭配是语言修辞的核心,词语搭配的发现对修辞表达以及从文本中挖掘信息等都有积极意义,而语境歧义词搭配的发现与提取对文本的情感倾向性分析有深刻意义。目前,带有情感的语境歧义词搭配的研究主要是针对某一领域,例如 Xiaowen Ding<sup>[3]</sup>在研究数码产品评论倾向性时考虑了语境歧义词,其主要是针对数码产品的属性、功能、外观等。针对语境歧义词给文本的情感分析所带来的问题,我们提出了“基于语境歧义词的句子情感倾向性分析”的方法,旨在为情感倾向性的判断提供基础和依据。

首先采用关联规则方法发现种子词的频繁集,通过 PMI 过滤机制消除无效关联项,剩余即为搭配词典候选集,对候选集进行情感倾向性判断,由有倾向性项构建成语境歧义词搭配词典。在倾向性分析时,句型分为转折句和非转折句,利用语境歧义词搭配词典,用语义分析方法完成情感倾向性判断。实验结果表明使用搭配词典后,准确率、召回率和 F-Score 都有一定的提高。

## 2 理论基础

### 2.1 关联规则在情感分析中的应用

关联规则是数据中蕴涵的一类重要规律,对关联规则进行挖掘是数据挖掘中的一项根本任务。Agrawal 等人<sup>[4]</sup>首次给出了两个项目集合(Item)中的元素在事务集合(Transaction)中存在的关联规则、置信度、支持度的定义。关联规则挖掘的经典应用是“购物篮”数据分析,目的是找出顾客在商场或者店铺所选购的商品之间的关联。下面给出一个关联规则的例子:

花生→啤酒 [支持度=40%, 置信度=80%]

这个规则表明 40%的客户同时购买花生和啤酒,在购买花生的人中 80%的人也购买了啤酒。

本文中,将关联规则应用在“如何发现语境歧义词搭配”问题上。词语的集合看成是两个 Item,词语搭配的集合看成是 Transaction,找出两个 Item 中的元素在 Transaction 上的并发关系。如下:

体重→暴涨 [支持度=20%, 置信度=70%]

夺取→生命 [支持度=20%, 置信度=70%]

这两个规则表明 20%的人同时使用“体重暴涨”和“夺取生命”,在分别使用“体重”和“夺取”的人中有 70%的人也分别使用了“暴涨”和“生命”。在“花生→啤酒”例子中没有考虑购买“花生”和“啤酒”的先后问题,在语境歧义词搭配中也是不考虑词语的先后问题,例如“视觉→对抗”和“对抗→政府”,表达了明显相反的两种情感倾向性,对于“对抗”,和其搭配的“视觉”和“政府”也没有先后的问题。下面给出两个定义,其中 X,Y 是 Item 中的两个元素:

定义 1 搭配集:形如  $X \rightarrow Y$  的规则如果满足最小支持度<sup>[5]</sup>( $\alpha$ ),称  $X \rightarrow Y$  构成的集合  $\{X,Y\}$  是搭配集。

定义 2 频繁集:搭配集中满足最小置信度<sup>[5]</sup>( $\beta$ )的规则  $X \rightarrow Y$  称作频繁集。

本文也采用类似关联规则的方法,在 Transaction 中先发现搭配集,然后在搭配集中识别频繁集,经过 PMI 过滤后构建语境歧义词候选搭配集,进而构建语境歧义词搭配词典。

### 2.2 基于语义分析的句子情感倾向性分析

目前,对于句子情感倾向性的分析,主要包含机器学习<sup>[6,7]</sup>和语义分析两类方法。

语义分析是用于知识的获取和展示的理论和方法<sup>[8]</sup>,它使用统计方法对语料集进行分析,提取和表示出词的语义,其是词语的上下文信息的总和。这是因为上下文对其中的事物提供了一组相互的联系和制约,这在很大程度上决定了词语之间语义上的相关性<sup>[8]</sup>。语义分析最初的应用是在信息检索中,解决关键字检索时的同义词和多义词问题。语义分析假设词语在文本中的使用模式存

在某种语义结构, 同义词之间具有基本相同的语义结构, 多义词的使用具有多种不同的语义结构, 语义分析的方法可以提取并量化这些语义结构, 进而消除语境歧义的影响, 克服多义词、同义词和单词依赖的现象, 提高文本表示的准确性。故使用语义分析方法进行句子的倾向性分析, 这种语义结构蕴含在语料中词语的上下文使用模式中。

### 3 语境歧义词搭配的认识

#### 3.1 种子词的选取

在 COAE2008 评测后, 分析结果发现判断错误的句子大多数为以下两种情况: (1) 句子中有语境歧义情感词, 无情感词, 如: “最近, 章子怡人气暴涨, 电影票房直线上升”。(2) 句子中不存在情感词也没有语境歧义情感词, 但却表达了情感, 如 “看着这个电影我就想睡觉!”。

为解决第一种类型的句子, 我们选择的种子词是 20 个常见的语境歧义词, 通过观察发现其他大部分的语境歧义词可以用 HowNet<sup>[9]</sup>和哈工大的同义词词林对这 20 个语境歧义词进行扩展得到, 这样对以后语境歧义词搭配词典的扩充将会带来方便。

#### 3.2 语境歧义词搭配的挖掘

对语料中的句子分词, 以种子词为中心词, 4 为窗口, 前后共取 8 个词。根据汉语语言思维, 修饰一个词语的词数在 1-6 个, 人们不倾向用长句, 长句会导致相互依存的两个词不再依存, 或不再属于同一个意群, 出现歧义。窗口取得太小会导致一些搭配未被发现, 过大会导致噪音过大, 故我们选择 4。把这 9 个词看作是一个序列串, 种子词组成的集合构成一个 Item, 里面的每一个元素为 X, 其他词组成的集合构成另一个 Item, 里面的每一个元素为 Y。蕴涵关系式  $X \rightarrow Y$  构成了搭配集。事务 Transaction 中的元素由每一个句子中的种子词分别和上下共 8 个词组合而成。在给定  $\alpha$  和  $\beta$  下, 采用如图 1 所示的流程挖掘语料中的搭配对。

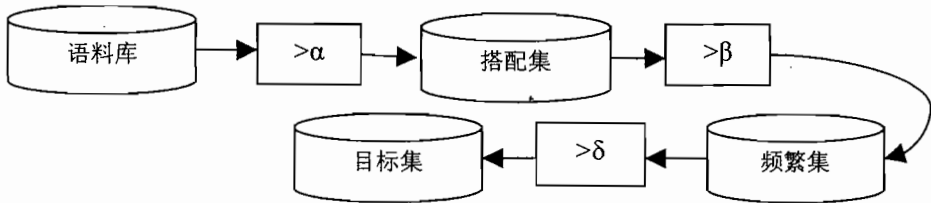


图 1 识别语境歧义词搭配

设  $\alpha=0.005\%$ 且以  $0.005\%$ 增长,  $\beta=0.1\%$ 且以  $0.1\%$ 增长, 针对每一对参数 $(\alpha, \beta)$ 训练, 取结果最好的一对。初始值的确定是因为当  $\alpha < 0.005\%$ ,  $\beta < 0.1\%$ 时剪枝过程不明显, 搭配集的噪音过大。

表 1  $\gamma$  的选取

$\alpha \backslash \gamma \backslash \beta$	0.00005	0.001	0.00015
0.001	0	0.03231	0.0677
0.002	0.0612	0.0513	0.0987
0.003	0.0738	0.0761	0.1111

在参数最优化时, 取多褒义的“变革”、多贬义的“暴涨”和褒贬个数相当的“减少”作为测试数据, 然后取三组实验的平均值。当两个参数增长时, 挖掘的搭配对逐渐减少, 带有情感倾向性的搭配对也随之流失, 用  $\gamma$  表示带有情感倾向性的搭配未被发现的比例。具体结果如上表 1 所示: 当  $\alpha=0.0001$ ,  $\beta=0.0001$  时,  $\gamma$  最小。

实验中通过观察生成的频繁集，我们发现很多生成的规则也是无关紧要的。基于这种问题我们利用信息论中的互信息来进一步挖掘搭配词间的关系。设两个词  $x$  与  $y$ ，出现的概率为  $P(x)$  与  $P(y)$ ，它们共同出现的概率为  $P(x,y)$ ，则互信息  $I(x,y)$  的定义如下：

$$I(x,y) = \log_2 \frac{p(x,y)}{p(x)p(y)} \quad (1)$$

设定一个阈值  $\delta$ ，如果  $I(x,y) < \delta$ ，认为该搭配无效。在公式(1)中，通常采用计算词的个数的方法计算概率。我们知道用词的  $tf-idf$  值计算互信息更能表达词与词的关联程度，所以把公式(1)改写成公式(2)。其中  $tf-idf_x$  为语料集中出现词  $x$  的  $tf-idf$  值。

$$I(x,y) = \frac{\sum tf-idf_x \times tf-idf_y}{\sum tf-idf_x \times \sum tf-idf_y} \quad (2)$$

与  $\alpha$ 、 $\beta$  的确定方法类似， $\delta < 0.02$  时，实验得到对频繁集的剪枝不明显，故初始  $\delta = 0.02$  且以 0.02 的速度增长，用三组实验的平均值寻找最优参数。用  $\mu$  带有情感倾向性的搭配流失的比例，最优化过程如表 2 所示， $\delta = 0.0309$  时实验结果最好。

表 2  $\mu$  的选取

$\delta$	0.02	0.04	0.06	0.08	0.1
$\mu$	0	0.0476	0.0309	0.0350	0.0427

对上述过滤后得到的目标集进行情感倾向性的判断，由带有情感倾向性的关联项成语境歧义词搭配词典，采用如下的存储形式：

<种子词 描述对象 情感类别> 或者 <描述对象 种子词 情感类别>

例如<颈椎 突出 负>，<成绩 突出 正>，<洪水 暴涨 负>，<人气 暴涨 正>。

## 4 句子的情感倾向性判断

本文在判断句子情感倾向性时采取语义分析方法。先对语料中能够体现情感色彩的词或者短语进行抽取，然后对抽取出来的词或者短语进行倾向性判断，并且赋予一个倾向值，最后将上述所有倾向值累加起来得到一个分值，根据这个分值来判断句子的情感倾向性。

### 4.1 基于语境歧义词的情感词典

本文采用大连理工大学信息检索实验室的情感词汇本体<sup>[10]</sup>、情感常识库<sup>[11]</sup>、基于第三部分建构的语境歧义词搭配词典和从 HowNet 的评价词表中抽取的评价词，共四部分构成本文的情感词典——基于语境歧义词的情感词典，以下简称“情感词典”。情感词典中词语的数目和正负项的分布如表 3 所示，其中 O 表示正、负项，D 表示情感词典的组成：

表 3 基于语境歧义词的情感词典

O \ D	情感词汇本体	情感常识库	搭配词典	HowNet	Total
正项	4917	5323	929	201	11370
负项	5342	871	698	167	7078

如果句子中有否定词，则句子的倾向性可能发生变化。本文为抽取的待判断词设置一个检测窗口  $winsize$ ，在  $winsize$  内出现否定词，就对句子的情感倾向性取反。根据实验测试得到当  $winsize$  为 6 时实验效果最好。相对与传统的否定词和转折词，在否定词表中加入了毫不、没法、难免、不怎么等重要的否定词。在转折词表中加入了不料、偏偏、谁知、岂知等有汉语特色的转折词。

## 4.2 句子倾向性判断

结合汉语思维和句法规则, 本文将句子分为转折句和非转折句。分词后判断句子中是否有转折词, 根据汉语思维习惯, 如果有, 认为句子要表达的中心思想在转折词后。首先为情感词汇表中情感词赋权重, 正项情感词+1, 负项情感词-1。对句子分词, 判断是否为情感词典里面的词, 然后在 *winsize* 内检测是否有否定词出现, 最后用公式(3)计算句子的情感分值。

$$\text{score} = \frac{\sum_{i=1}^m p_i d_i + \sum_{j=1}^n n_j d_j}{m + n} \quad (3)$$

其中  $p_i=1$  代表正项情感词权重,  $n_j = -1$  代表负向情感词权重, 当正项词或者负项词的前面存在否定词时,  $d_j = -1$ 。  $m$  和  $n$  分别代表句子中情感词存在的个数。当 *score* 大于 0 时句子的情感倾向性即为正项, 当分值小于 0 时句子的情感倾向性即为负项。

## 5 实验结果与分析

首先我们用第四节中提到的方法对句子进行情感倾向性分析, 为了证明语境歧义词搭配词典的重要性, 我们把情感词典分成两部分: 一部分不含有语境歧义词搭配词典; 另一部分含有语境歧义词搭配词典。

我们首先选择的是 COAE2008 的语料。因为 COAE2008 语料涵盖的能容比较全面, 其中包括科技、财经、体育、电影影评、新闻等各个方面的内容, 较具有空间跨度, 覆盖面比较广, 不单一。为了验证所建立的情感歧义词搭配词典是否有通用性, 即不只是对 COAE 语料有提高, 我们又在大连理工大学信息检索实验室的情感语料库<sup>[12]</sup>上作了对比试验, 情感语料库中包含 250021 个句子, 句子覆盖小学教材、电影剧本、童话故事、文学期刊。从时间、空间、学科、风格和构成上看覆盖面大。其中包含本文所选的种子词但不包含情感词汇本体中的词的句子占 5.8%。

### 5.1 语料选取与预处理

首先将两个语料中的短篇篇文章进行分句, 检索出其中带有种子词的句子, 每个种子词取出其中的主观句子 500 条作为实验语料。在选择实验语料时要剔除存在以下三种情况的句子:

- (1) 经过分词后不再包含种子词的句子。
- (2) 不成句, 不能表达完整意思的句子。
- (3) 与同一种子词的已有句子表达意思相同的句子。

其中, 条件 (1) (2) 用来保证所选语料真正包含所选词汇并且符合汉语语法的完整句子, 这个问题的产生是由于在语料中存在大量的网络上关于电影、体育的评论, 而网评信息的一个特点就是语法不规范, 存在较多的语病, 这会影响本文对语境歧义词的研究; 条件 (3) 是为了排除语料中大量重复信息带来的干扰。句子中存在部分相同的成分时, 我们视为是该词汇的不同句法或者正常的语用重现, 直接保留不予剔除。句子数 500 条是人工观察选定的, 因为超出 500 后所获取的重复句子明显增加, 很少能获得新的实例。语料收集完成后, 人工对语料进行标注, 若句子表达的是褒义则标注为“正”项, 若句子表达的是贬义则标注为“负”项。20 个种子词语在语料中的“正”项、“负”项以及每个种子词中所包含的隐式情感句的个数分布情况如图 2 所示: 其中 724 个隐式情感句子中, 正项隐式情感句有“289”个, 负项隐式情感句有“435”个。

句子数 500 条是人工观察选定, 超出 500 后获取的重复句子明显增加, 很少能获得新的实例。人工对语料标注为“正”项和“负”项。正项情感句 5574 个, 负项情感句 4426 个, 724 个隐式情感句子中, 正项隐式情感句有“289”个, 负项隐式情感句有“435”个。

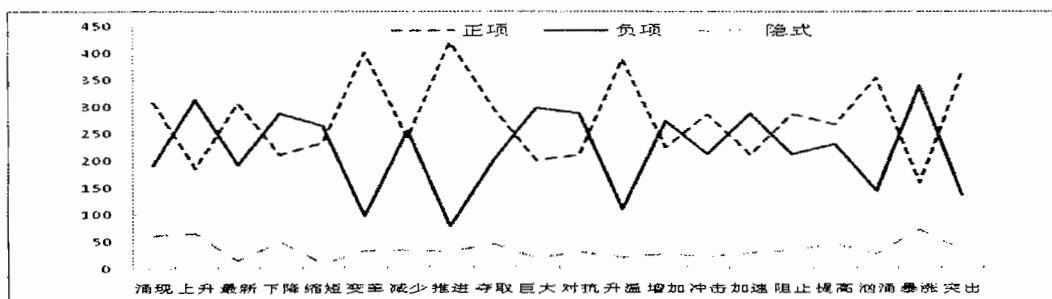


图2 语料分布情况

## 5.2 实验结果与分析

在 COAE2008 中和 20 个种子词为类别的实验结果总体上都有一定的提高,从表 4 中可以看到“暴涨”、“上升”、“涌现”的准确率、召回率和 F 值相对提高的多,而“最新”、“缩短”、“升温”、“巨大”的准确率、召回率和 F 值相对提高的少。通过观察语料,发现提高得多的种子词在语境歧义词搭配词典中所占的比例要比其他的种子词所占的比例大,这类词在语料中存在语境歧义的情况较明显。两个形容词性的种子词的百分率都提高的不多,这也正符合我们的汉语语言习惯。我们知道当形容词修饰名词时,情感的倾向性或者体现在名词上面或者体现在形容词上面,看下面的两个例子:(1)“美丽的女孩”和“悲伤的事情”的情感倾向性主要体现在形容词“美丽”和“悲伤”上。(2)“巨大的灾难”、“最新疫情”的情感倾向性体现在名词“灾难”和“疫情”上,而类似“灾难”和“疫情”这些带有情感倾向性的名词已经收录在我们的情感词汇本体中,所以提高的不多。由此我们也得出,带有情感倾向性的语境歧义词主要是动词。

表4 20个种子词的实验结果

种子词	不含搭配词典			包含搭配词典		
	Precision	Recall	F-Score	Precision	Recall	F-Score
上升	0.7923	0.7681	0.7800	0.8821	0.8695	0.8775
下降	0.8126	0.8136	0.8131	0.8763	0.8751	0.8757
夺取	0.7652	0.8232	0.7931	0.8731	0.8931	0.883
汹涌	0.7531	0.7914	0.7717	0.8425	0.8594	0.8508
巨大	0.7563	0.6974	0.7257	0.7698	0.7098	0.7386
变革	0.7295	0.7412	0.7353	0.7502	0.7563	0.7532
加速	0.7101	0.7448	0.727	0.7312	0.7501	0.7405
阻止	0.6987	0.7145	0.7065	0.7128	0.7325	0.7225
提高	0.7251	0.6857	0.7048	0.7612	0.7011	0.7299
突出	0.6942	0.7471	0.7197	0.7421	0.7622	0.752
增加	0.7024	0.6855	0.6938	0.7211	0.6994	0.7101
减少	0.7498	0.6854	0.7162	0.7523	0.7264	0.7391
推进	0.6875	0.7314	0.7088	0.7211	0.7541	0.7372
对抗	0.6788	0.7114	0.6947	0.7241	0.7456	0.7347
暴涨	0.7814	0.8245	0.8023	0.8934	0.8617	0.8773
最新	0.7734	0.7412	0.7569	0.7902	0.7562	0.7728
缩短	0.6715	0.7058	0.6882	0.6821	0.7221	0.7015
升温	0.7864	0.7321	0.7583	0.8011	0.7531	0.7763
汹涌	0.6917	0.7213	0.7062	0.7226	0.7403	0.7313
冲击	0.7514	0.7021	0.7260	0.7695	0.7436	0.7563

我们又分别对 COAE2008 语料中的正、负例情感分别做了对比试验，试验结果如下表 5。

表 5 COAE2008 语料中正、负例的实验结果

含搭配词典否	正例			负例		
	Precision	Recall	F-Score	Precision	Recall	F-Score
不含	0.7963	0.8124	0.8024	0.7762	0.8022	0.789
含	0.8302	0.8325	0.8313	0.7933	0.8119	0.8025

我们发现正类的分类效果优于负类，仔细阅读语料发现，负类中隐性情感的比例要多于正类，并且负类中对于情感的表达方式比较复杂，结构也比较多样化。观察正负例分类出错的句子，出错原因主要有以下几点：（1）长句占据的比例比较多，并且这类句子一般含有多个情感主题，没有一个主要的情感倾向性。（2）由于用支持度、置信度、PMI 过滤导致一小部分的带有情感倾向性的搭配对流失，使得判断错误，但是由流失的搭配引起的分错类别的句子占极少数。（3）对于句子中不存在情感词也没有语境歧义情感词，但是却表达了一定的情感的问题，如“看着这个电影我就想睡觉！”，这种句子我们还没有很好的解决，而这也一直是句子情感倾向性判断的一大难题。

为证明语境歧义词搭配词典的通用性，我们在 COAE2008 和情感语料库上分别做了对比试验，分别如图 3、4 所示，从图中看到两者的实验结果都有一定提高。分析所有的实验结果，我们认为，提高的原因在于：基于语境歧义词搭配词典的实验不仅考虑了句式，句法规则，而且还充分考虑了语境歧义词、否定副词和转折词对语言情感倾向性的影响。

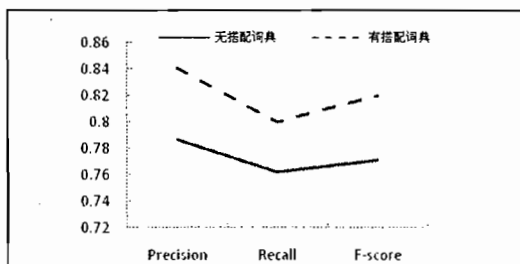


图 3 COAE2008 语料集的实验结果

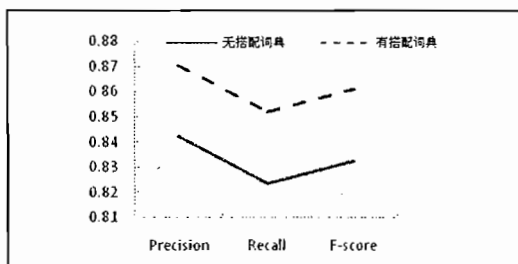


图 4 情感语料库的实验结果

## 6 结语

本文给出了一种基于关联规则挖掘语境歧义词搭配频繁集，构建语境歧义词搭配词典的方法。从语义搭配的角度出发，把句子分为两大类，设计了两种观点倾向性分析策略。对语料的测试表明，基于语境歧义词的句子倾向性分析有重要的意义，我们从汉语语义上的修饰关系入手，挖掘情感评价对象和评价词之间的深层联系。通过在 COAE2008 语料上和情感语料库上的实验，我们发现了语境歧义词搭配词典在文本倾向性分析中的重要性。

目前语境歧义词搭配词典的建设工作还在进行中，目前，其中只含有与种子词和与种子词意思相近或者相反的搭配。我们计划加入更多的语义资源来丰富语境歧义词搭配词典，根据实际的需要增加语境歧义词搭配词典的规模，使语境歧义词搭配词典更加的全面、完善。

## 参考文献

- [1] Bolshakov L.A. Gelbukh A. Heuristics-Based Peplenishment of Collocation Databases[J]. E.M.Rancbhod and N. J. Mamede(E)ds. 2002(2389): 679-707.
- [2] 车万翔, 刘挺, 秦兵. 面向依存文法分析的搭配抽取方法研究[C]. 全国第六届计算语言学联合学术会议, 太原, 山西. 2001: 102-107.

- [3] X Ding, B. Liu, P S Yu. A Holistic Lexicon-Based Approach To Opinion Mining[C]. In The Proceedings Of The Conference On Web Search And Web Data Mining (WSDM).Palo Alto, California, United States. 2008: 231-240.
- [4] R.Agrawal, T.Imielinski, A.N.Swami. Mining Association Rules Between Sets Of Items In Large Data Bases[C]. In The Proceedings Of The ACM SIGMOD Intl Conference On Management Of Data. Washington, USA. 1993: 207-216.
- [5] R.Agrawal, R.Srikant. Fast Algorithms For Mining Association Rules[C]. In The Proceedings Of Intel Conference On Very Large Data Bases. Santiago, Chile.1994: 487-499.
- [6] Soo-Min Kim, Eduard Hovy. Automatic Detection Of Opinion Bearing Words And Sentences[C].In The Proceedings of IJCNLP-2005.JeJu Island, Korea. 2005: 61-66.
- [7] Jun Zhao, Kang Liu. Adding redundant features for crfs-based sentence sentiment classification[C].In The Proceedings Of The 2008 Conference On Empirical Methods In Natural Language Processing. Honolulu, Hawaii. 2008: 117-126.
- [8] Thomas K. Landauer, Peter W. Foltz, Darrell Laham . Introduction to Latent Semantic Analysis Discourse Processes[M], 1998. 25(2): 259-284.
- [9] Zhen Dong. Dong.[http://www.keenage.com/zhiwang/e\\_zhiwang.html](http://www.keenage.com/zhiwang/e_zhiwang.html).
- [10] 徐琳宏, 林鸿飞, 潘宇. 情感词汇本体的构造[J]. 情报学报. 2008. 27(2): 180-185.
- [11] 陈建美, 林鸿飞. 中文情感常识知识库的构建[J]. 情报学报. 2009. 28(4): 492-498.
- [12] 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): 116-1.