

文章编号:1003-0077(2006)01-0014-07

## 基于 HowNet 的词汇语义倾向计算\*

朱嫣岚, 闵 锦, 周雅倩, 黄萱菁, 吴立德

(复旦大学 计算机科学与工程系, 上海 200433)

**摘要:**在互联网技术快速发展、网络信息爆炸的今天,通过计算机自动分析大规模文本中的态度倾向信息的技术,在企业商业智能系统、政府舆情分析等诸多领域有着广阔的应用空间和发展前景。同时,语义褒贬倾向研究也为文本分类、自动文摘、文本过滤等自然语言处理的研究提供了新的思路 and 手段。篇章语义倾向研究的基础工作是对词汇的褒贬倾向判别。本文基于 HowNet,提出了两种词汇语义倾向性计算的方法:基于语义相似度的方法和基于语义相关场的方法。实验表明,本文的方法在汉语常用词中的效果较好,词频加权后的判别准确率可达 80% 以上,具有一定的实用价值。

**关键词:** 计算机应用; 中文信息处理; 态度分类; 语义倾向; 知网

**中图分类号:** TP391      **文献标识码:** A

## Semantic Orientation Computing Based on HowNet

ZHU Yan-lan, MIN Jin, ZHOU Ya-qian, HUANG Xuan-jing, WU Li-de

(Department of Computer Science and Technology, Fudan University, Shanghai 200433, China)

**Abstract:** Nowadays, with the development of Internet and information explosion, automated techniques for analyzing author's attitudes towards specific events will make great effort to business intelligence and public opinion survey. Semantic orientation inference has become a meaningful tool, which could provide useful information for text classification, summarization, filtering etc. Measuring the semantic orientation of words would greatly contribute to predicting the author's attitude in a passage. In this paper, a simple HowNet-based method for semantic orientation computation of Chinese words is introduced. Although this method requires only a few seed words, satisfactory results can still be obtained. And the performance is even better for frequently used words, with the frequency-weighted accuracy of above 80%.

**Key words:** computer application; Chinese information processing; sentiment classification; semantic orientation; HowNet

## 1 简介

面对浩瀚的网络资源,如何通过有效的手段获取所需的信息,是目前计算机领域研究的热点。

语义倾向性研究正是在这一背景下应运而生的。人们根据各自所关注的问题在 Web 上进行搜索,通过阅读大量相关文献、介绍、评论,逐渐形成自己的观点。由于网络资源过于庞大,搜集-阅读-评价的工作变得十分繁琐和耗时。因此迫切地需要一套能够自动搜索、评价、归纳的系统替人们完成这一任务。困难是显而易见的,计算机虽然具有极快的处理速度,

\* 收稿日期:2005-05-07 定稿日期:2005-11-01

基金项目:国家自然科学基金资助项目(60435020);上海市科技攻关计划资助项目(035115028)

作者简介:朱嫣岚(1982—),女,硕士生,主要研究方向为自然语言处理。

但如何使其能够自动完成对文本的语义倾向的推断,仍然是个需要研究的课题。

所谓词汇的语义倾向,即对于词汇的褒贬程度计算出一个度量值。为了便于统计和比较,目前比较常用的做法是将度量值规定为位于  $\pm 1$  之间的实数。当度量值高于某阈值时,判别为褒义倾向;反之,则判为贬义倾向。此外,我们可以通过对篇章中词汇的语义倾向值求平均的方式,获得篇章的语义倾向。因此,对词汇的语义倾向计算是此类研究中的关键工作。

语义倾向研究具有极大的实用价值。在商业领域中,能够为企业进行市场分析、市场调查、顾客反馈提供更多有价值的信息;在管理领域,能够帮助领导者更快地了解群众对各类政策措施的反馈意见。

此外,语义倾向判别也为文本过滤、自动文摘的研究工作提供了新的思路和新的手段。我们可以对语义倾向度量值设定一个合适的阈值,对于倾向值低于或高于阈值,也就是态度倾向过于偏激的文章进行过滤操作,或者可将倾向值赋予一定的权值,作为文本过滤中需要考虑的一个因素。该方法如果应用在网络新闻组的自动过滤中,有利于维护讨论秩序,缓和讨论气氛。而自动文摘的生成,可将具有强烈褒贬倾向的语句作为关键句摘出,从而更好地保留原作者的观点和意见。

单词的语义倾向判别是篇章语义倾向研究的基础。早在 1997 年,Hatzivassiloglou 和 McKeown 就尝试使用监督学习的方法对词语进行语义倾向判别,通过对训练语料的学习进行语义倾向判别,准确率约 82%,在加入篇章中形容词之间的接续信息后,准确率提升到 90% 左右<sup>[1]</sup>。在 2003 年,Turney 在其论文<sup>[2]</sup>中提出了利用统计信息对单词进行语义倾向判断的新方法。其准确率在包含形容词、副词、名词、动词的完整测试集上达到 82.8%。

文本的语义倾向判别也可被看作一个褒贬的分类问题,因此,文本分类中的方法同样被应用到了语义倾向判别研究中。2002 年,Turney 在其论文<sup>[3]</sup>中介绍了基于语义倾向的非监督文本分类方法。根据褒贬含义的倾向信息对评论性文章进行分类,其分类结果更符合人们对评论性文章分类的实际需求。在 Epinions 上的 410 篇评论性文章中,利用 Turney 的算法对褒贬倾向进行分类,正确率达到 74%。同年,Pang 等人使用标准的机器学习技术做了同样文本分类工作<sup>[4]</sup>。比较了简单贝叶斯、最大熵、SVM 方法在语义倾向文本分类上的效果。SVM 的分类准确率最高达到约 80%,为几种方法中分类效果最好的。之后,Pang 提出了对语义倾向分类结果进行后续处理,进一步将分类结果按照语义倾向强度进行细分的方法,并通过一系列的实验证明了其方法的可行性<sup>[5]</sup>。

近年来,陆续出现了一些利用语义倾向分析技术开发的商业智能系统,例如,NEC 公司的 Kusha 等人所开发的 ReviewSeer,通过对评论性文章的语义倾向分析,为商品的受欢迎程度进行打分评价,该评价结果是极具价值的商业信息<sup>[6]</sup>。又如,Bing Liu 在其论文<sup>[7]</sup>中介绍了商用产品信息反馈系统 Opinion Observer,利用网络上丰富的顾客评论资源,进行商品的市场反馈分析,为生产商和消费者提供了直观的针对商品各个特性的网络评价报告。

本文的研究侧重于词汇的语义倾向性判别。利用 HowNet<sup>[8]</sup>提供的语义相似度和语义相关场计算功能,尝试了一些词语语义倾向判别的方法。主要做了以下工作:通过计算词语间的相似程度,对词语的褒贬倾向按照一定的计算法则进行赋值。根据所得的语义倾向度量值判别其褒贬倾向。

本文其他部分的组织如下:

第二节对 HowNet(《知网》)进行简单介绍,并且提出了两种基于 HowNet 的词汇倾向性计算方法。

第三节对使用本文算法所进行的实验进行详细描述,对实验的结果进行分析。

第四节对于本文所做的工作做了一个总结,并提出了未来的研究方向。

## 2 词汇的语义倾向计算

### 2.1 HowNet(《知网》)介绍

知网(英文名称为 HowNet)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>[8]</sup>。

对于汉语词汇,知网中的描述基于“义原”这一基本概念。义原,可以被认为是汉语中最基本的、不易于再分隔的最小语义单位。由于汉语中“词”的含义非常复杂,往往一个词在不同的语境中会表达不同的语义。因此,在 HowNet 中,把汉语中的词理解为若干义项的集合。《知网》的语义字典中,每条记录都是由一个词语的一条义项及其描述所组成,即一条记录对应与一个词语的一个义项。

本文的算法分别使用了 HowNet 提供的语义相似度和语义相关场的计算功能。语义相似度的计算功能是根据刘群的论文<sup>[9]</sup>中的原理编写的词汇语义相似度计算程序,实现了义原之间语义相似度的计算。通过输入两个词语并分别选取确切的义原,在结果显示框中即可得到相似度的数值。举个简单的例子,输入词语“好”、“差”,并分别选择相应的义原为“{ HighQuality | 优质}”、“{ unqualified | 不合格}”,得到的输出结果为 0.021053,即这两个词语在相应义原下的相似度。语义相似度主要反映的是词语语义的相似程度。在机器翻译中,可以理解为两个词语在不同的上下文中可以互相替换使用而不改变文本的句法语义结构的程度。HowNet 相似度值被定义为 0 到 1 之间的一个实数。至于词与词之间的语义相似度,我们使用了词语两两义原相似度的最大值,希望最大限度的保留词语之间相似性信息。

所谓语义相关场,不同于语义相似度,反映的是两个词语互相关联的程度。可以用这两个词语在同一个语境中共现的可能性来衡量。因此相关的词语语义上可能相近,也可能相反。通过输入一个词语并选取词语的某一义原,可以获得与该义原语义相关的一组词,例如,对于“漂亮”这个词,若选择其义原为“{ beautiful | 美}”,即可获得一组语义相关的词语,包括“美貌”、“美景”、“俏丽”、“美丽”、“好看”等。同时语义相关场的范围(即语义相关词语的数量)可通过参数调整(分三级,从小到大分别是 Rank1 ~ Rank3)。

本文的算法利用了 HowNet 的语义相似度和语义相关场功能,希望能通过其所提供的语义信息为词语褒贬判别找出一条新的途径。

### 2.2 原理

我们希望为每个单词赋予一个语义倾向的度量值。其大小由这个单词与基准词的语义关联的紧密程度有关。基准词,在这里指褒贬态度非常明显、强烈,具有代表性的词语。与褒义基准词联系越紧密,则词语的褒义倾向越强烈。与贬义基准词联系越紧密,则词语贬义倾向越明显。基于这样的设想,设计如下实验:

假设共有  $k$  对基准词,每对基准词包括一个褒义词和一个贬义词。褒义基准词表示为  $key\_p$ ,贬义基准词表示为  $key\_n$ ,单词  $w$  的语义倾向值用  $Orientation(w)$  表示,以 0 作为默认阈值,最终倾向值大于阈值为褒义,小于阈值为贬义。 $Orientation(w)$  数值大小代表单词  $w$  褒贬强烈程度。

单词  $w$  的语义倾向值计算公式如下:

$$Orientation(w) = \sum_{i=1}^k Similarity(key-p_i, w) - \sum_{j=1}^k Similarity(key-n_j, w) \quad (1)$$

我们的试验中,公式(1)中的  $Similarity(key, w)$  的计算分别采用了 HowNet 语义相似度和语义相关场两种方法。

#### 1. 利用 HowNet 语义相似度计算单词的语义倾向值

$$\text{公式(1) 中的 } Similarity(key, w) = Sim(key, w) \quad (2)$$

利用 HowNet 计算两个单词义原之间的相似度,取两个单词义原之间相似度的最大值作为  $Sim(key, w)$ 。

#### 2. 利用 HowNet 语义相关场计算单词的语义倾向值

$$\text{公式(1) 中的 } Similarity(key, w) = \frac{|Relevance(w) \cap Relevance(key)|}{|Relevance(w) \cup Relevance(key)|} \quad (3)$$

HowNet 的语义相关场的获得是以义原为单位,即给出一个义原,可获得属于其相关场的单词集合。这里所说的  $Relevance(w)$  表示单词  $w$  所有义原的语义相关场词语的并集。式(3)的分子和分母分别表示两个词语相关场交集和并集的元素个数。

### 3 实验

#### 3.1 测试词表

度量语义倾向判别效果的方法有两种,一是直接判断其算法对单词的褒贬倾向判断的准确率,二是通过语义倾向判别为某个实际应用带来的性能提高来衡量。本文使用前一种较直观的方式对实验效果进行评价。

语义倾向实验所使用的测试集共有 3 组。测试集 1 ~ 测试集 3。

表 1 测试集

测试集编号	褒义词数	贬义词数	总词数
1	3146	3299	6445
2	1370	1425	2795
3	69	211	280

测试集 1 使用了 HowNet 第一版中文词表中标注“良”(褒义)、“莠”(贬义)属性的词汇。总共选用 6445 词。其中褒义词 3146 个,贬义词 3299 个。

在实验过程中,我们发现,测试集 1 中包含这样一些词语,其义原集合中某一义项虽然被标注为褒义或贬义,但该义项作为该词的一个特殊义项并不常用;或者该词语在不同语境下,或为褒义,或为贬义,并不能简单地将其判断为褒义词或贬义词。例如:词语“好看”,在描述事物时,可作褒义,如“这花真好看”,而在“要你好看”这样的语句中,显然带有强烈的贬义。为了使实验结果更能说明问题,更具有说服力,我们从测试集 1 中人工挑选了 2795 个褒贬含义较明确的词语作为测试集 2。

为了测试我们的方法在常用词中的效果,我们把测试集 2 中的词语按 Google 搜索返回 Hits 数(截止至 2004 年 8 月 31 日),即它们在 Web 上的词频降序排列,选取词频最高的 280 个词语,作为测试集 3。

表 1 中列出了各个测试集合的具体信息。

#### 3.2 基准词选择

在我们的方法中,计算语义倾向度必须首先选择一组褒贬基准词对,根据被测试词与这组基准词的语义关系紧密程度,通过计算得到语义倾向值。显然,所选择的基准词必须是具有强

烈褒贬倾向,并且具有代表性的词语。基于以上原因,我们选择了测试集 2 中的词语作为基准词对的候选集合,根据词频对词语进行降序排列,分别从褒贬词中选取出现频率最高的作为褒贬基准词。总共选择了褒贬词对 3 组,分别进行实验比较。基准词对 1 由词频最高的前 10 个褒义词和前 10 个贬义词组成,基准词对 2 选取了词频最高的前 20 对褒贬基准词,基准词对 3 选取了词频最高的前 40 对褒贬基准词。表 2 给出了基准词对 3 的全部词汇,而基准词对 1 和基准词对 2 显然都是它的子集。本文所有实验结果都是从测试集中扣除了这 40 组基准词后得到的。

表 2 40 组褒贬基准词

褒义基准词:

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	简单	和平	开通	真实
先进	便宜	优质	欢乐	美好	良好	不错	出色	成熟	完善

贬义基准词:

不合作	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

### 3.3 实验概述

实验的目的是为每个测试集中的单词赋予语义倾向值,通过倾向值来度量其褒贬程度。在以下实验中,默认使用 0 为阈值,即倾向值大于 0 则判断为褒义,小于 0 则判断为贬义。语义倾向判别准确率 = 判别正确的词数/测试集总词数,我们以此来衡量算法效果。

语义倾向值的计算使用了两种方法,分别利用了 HowNet 的语义相似度和语义相关场的计算功能。具体计算公式在 2.2 中已有详细介绍。

首先,对于这两种计算方法的效果进行了实验,在此,我们使用随机赋值的方法作为基准参与比较。为了避免获得的相关场集合较小而导致  $Relevance(w) \cap Relevance(key)$  交集为空的情况,因此,我们选用可调参数  $Rank = 3$ ,获得单词的最大的语义相关场。

实验使用基准词对 3(40 对褒贬基准词),实验结果如下表:

表 3 Similarity(w) 在不同计算方法下的语义判别准确率(使用 40 对基准词)

测试集编号	1	2	3
HowNet 相似度计算	68.04 %	68.32 %	85.50 %
HowNet 相关场计算	62.03 %	63.43 %	75.00 %
随机赋值(baseline)	49.74 %	51.16 %	45.00 %

表 3 显示了在 3 个测试集上,两种不同的计算方法对语义倾向判别中的效果。我们发现,在同一测试集上,使用相同的基准词对,基于 HowNet 语义相似度的计算方法比基于语义相关场的计算方法准确率高。其原因主要是由于,语义相关场的定义中,所谓相关,既包括了褒贬倾向相近的词语,同时又包括了褒贬倾向相反的一些词语,例如:“好”这个词,选择其义原为“优质”,所获得的相关场词集中,既包括了“优良”、“良好”这类语义相近的褒义词,又包含了“坏”、“劣质”这些语义相反的贬义词。由于这个原因,导致了语义相关场并不能将词语间褒贬的相近程度很好地反映出来。而语义相似度恰恰能在一定程度上反映词语间的语义倾向的一致性,因此实验效果较好。

为了更好地反映本文所提出的语义倾向算法在实际应用中的效果,我们把词语使用频率作为衡量算法性能的一个指标。我们对测试集中的单词按 Google 搜索返回 Hits 数进行了统计。由此作为该单词使用频率的度量指标。测试结果的准确率根据单词使用频率加权计算。

准确率 =  $\frac{\sum result(nw_i)}{\sum result(w_i)}$ , 公式中  $result(word)$  表示词语  $word$  的词频,  $nw$  代表褒贬倾向判断正确的单词,  $w$  表示测试集中的任意单词。

下表将经过词频加权的实验结果与初始实验结果在测试集 2 上进行了比较。语义倾向值的计算使用基于 HowNet 语义相似度的计算方法。

表 4 对准确率进行词频加权后的实验结果(测试集 2,使用 Hownet 语义相似度计算)

基准词对数	10	20	40
未经加权的结果	55.73 %	59.48 %	68.32 %
经过词频加权后的结果	60.77 %	64.22 %	78.58 %

由表 4 可以很明显地看出,根据单词词频加权的准确率较初始结果好得多。在不同基准词对下,其准确率高出 5 % ~ 10 %。由此说明,本文的算法在常用词中的判别效果较低频词好,因此本方法也更具有实际应用的价值。

另外,基准词对数目的大小对实验效果的影响也是我们所关心的一个问题。在测试集 2 中,我们挑选了使用频率最高的一部分词,作为测试集 3,而基准词仍然根据词频选取前 1 对、4 对、5 对、10 对、15 对、20 对、30 对、40 对褒贬词进行测试。词语相似度的计算仍使用基于 HowNet 语义相似度的方法。实验结果如下:

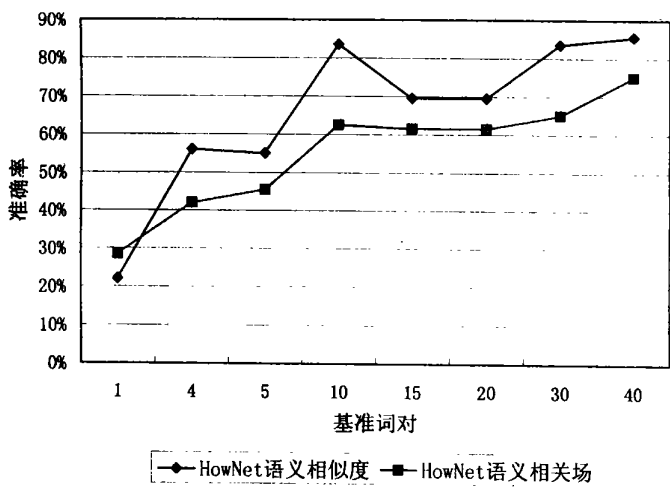


图 1 基准词数量对实验效果的影响(测试集 3)

从上图可以看到,语义倾向判别的准确率大致随基准词的数量增加而上升,也就是说,给出的褒贬“范例”- 基准词越多,算法的效果就越好。同时,与表 3 列出的结果相一致的,在测试集 3 上仍能得到相似度计算方法较相关场计算方法更优的结论,但随着基准词的数量增加,两种方法的差异逐渐减小。

以上的实验都将 0 作为默认阈值,计算所得的倾向值大于 0 的判为褒义词,小于 0 的则判为贬义词。考虑到基准词选择中,可能出现的偏向,作者尝试调整阈值并对实验结果进行了一系列分析,结果表明,在各个测试集上,阈值的调整对准确率的提高并不具有显著的作用,因此

在这里不做详细介绍。

## 4 结论

本文所提出的方法是作者在汉语语义倾向判别研究中的一个初步尝试。基于 HowNet 的语义倾向判别,只需利用 HowNet 的本地资源和少量的基准词,比较容易实现且不受外界条件(如网络环境)的干扰。从实验结果来看,基准词的增加对判别的准确性提高有明显作用,但即使是极少量基准词,在常用词集中同样可以达到 80 % 以上的准确率。

文中所使用的算法较简单,仍有许多值得商榷和改进的余地。例如,在基准词的选择方面,选择一些具有代表性且倾向强烈的基准词,对实验效果的提高将会有很大的帮助,作者的实验所使用的基准词只是简单的根据词频进行挑选,难免有一些偏颇和失当,在使用的算法方面,对于语义倾向值的计算方式比较直观,如果改用其他更加科学的计算方式,相信准确率会有更大的提高。作为篇章语义倾向判断的关键问题,本文在词语的褒贬倾向判别的方法上进行了一些探讨,期望对相关研究工作有所帮助。

## 参 考 文 献:

- [1] Vasileios Hatzivassiloglou, Kathleen R. McKeown. Predicting the semantic orientation of adjectives[A]. In: Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics and the 8th Conference of the European Chapter of the ACL[C], 1997:174 - 181.
- [2] Turney, Peter, Littman Michael. Measuring praise and criticism: Inference of semantic orientation from association[J]. ACM Transactions on Information Systems, 2003, 21(4): 315 - 346.
- [3] Turney Peter. Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews[A]. In: Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics[C]. 2002: 417 - 424.
- [4] Bo Pang, Lillian Lee, Shivakumar Vaithyanathan. Thumbs up? Sentiment classification using machine learning techniques[A]. In Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing[C]. 2002: 79 - 86.
- [5] Bo Pang, Lillian Lee. Seeing Stars: Exploiting Class Relationships for Sentiment Categorization with respect to Rating Scales[A]. ACL2005, 115 - 124.
- [6] K Dave, S Lawrence, DM Pennock. Mining the peanut gallery: opinion extraction and semantic classification of product reviews[A]. WWW2003, 519 - 28.
- [7] Bing Liu, Mingqing Hu, Junsheng Cheng. Opinion observer: analyzing and comparing opinions on the Web[A]. WWW2005, 324 - 351.
- [8] HowNet[R]. HowNet's Home Page. <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[A].第三届汉语词汇语义学研讨会,台北,2002.