第43卷 第6A期
2016年6月

计 算 机 科 学
Computer Science

Vol. 43 No. 6A
June 2016

VDEA词典的构建及其在情感倾向性分析中的应用

黄金柱[1] 李 峰[2,3] 张克亮[1]

(中国人民解放军外国语学院语言工程系 洛阳 471003)[1]

(北京航空航天大学软件开发环境国家重点实验室 北京 100191)[2]

(中国人民解放军后勤科学研究所 北京 100166)[3]

摘 要 配价语法主要以谓词为中心研究句子的深层语义结构,重点描述动词和形容词与搭配成分间的依存关系,是解决语义分析处理这个颈瓶问题的利器。以英语形容词为主体,构建了包含相关配价信息的形容词配价词典,词典包含3170个英语形容词的配价关系、格关系、释义、褒贬义、语义分类、语义特征和相关例句等信息。此外,基于该词典设计了词汇情感倾向性分析模型,取得了很好的实验结果。

关键词 配价语法,形容词,知识库,情感倾向性

中图法分类号 TP391 **文献标识码** A

Construction of VDEA and its Application in Lexical Sentimental Orientation Analysis

HUANG Jin-zhu[1] LI Feng[2,3] ZHANG Ke-liang[1]

(Department of Language Engineering, PLA University of Foreign Languages, Luoyang 471003, China)[1]

(State Key Laboratory of Software Development Environment, Beihang University, Beijing 100191, China)[2]

(Logistics Science Research Institute of PLA, Beijing 100166, China)[3]

Abstract Valency grammar focuses mainly on studying the deep semantic structure of sentences through analyzing predicates and attaches emphasis upon the dependent relationships between predicates and cooperative elements. Now the grammar is an edged tool to solve the problem of semantic analysis. The valency-based dictionary of English adjective (or VDEA) constructed in this study is a machine-readable dictionary based on valency grammar. VDEA covers 3170 English adjectives and relevant semantic information such as valency relationships, case relationships, explanation, commendatory and derogatory meanings, semantic classification, semantic features and examples etc. Based on the dictionary, the study designed a lexical sentimental orientation analysis model. The experimental result is satisfactory.

Keywords Valency grammar, Adjective, Knowledge base, Sentimental orientation

1 引言

20世纪50年代,法国语言学家Tesnière在《结构句法基础》(Elements De Syntaxe Structurale)中把"从属"这个概念第一次系统地运用到句法理论的各个方面,他被认为是配价语法理论的创始者。配价语法起源于法国,发展于德国,随后世界各国纷纷将其用于本国语言的研究。

自朱德熙于1978年引进配价语法以来,国内就掀起了对配价语法的研究热潮,主要代表人物包括朱德熙、袁毓林、邵敬敏、张国宪、沈阳、郑定欧、周国光等学者。受国外配价词典建设的引导和启发,越来越多国内学者也开始着眼于基于配价的语义知识库建设,其分为人用语义词典和机用语义词典两类,主要包括天津外语学院德语系与曼海姆德语研究所联合编写的《德汉动词配价对比词典》,这部词典以《德语动词配价词典》为基础,结合汉语配价研究成果编撰而成;此外还有北京大学的VCSD词典等[1]。随着语义研究的不断深入,配价语义资源建设的不足仍然是制约语言信息处理的颈瓶。

2 VDEA词典的设计与实现[1]

作为修饰性词汇,形容词语义资源在语言交流及信息处理中发挥着重要作用,对形容词的准确理解将直接影响到其它词汇乃至句子和篇章的理解。结合英语形容词自身特点并参阅相关知识库的语义描述方法,本研究构建了英语形容词配价词典(Valency-based Dictionary of English Adjective, VDEA)。VDEA是一个基于规则(rule-based)的语义词典,在进行形容词的语义信息描述过程中,结合了配价语法、格语法以及语义特征等语法理论,同时参考了论旨理论、语义网络以及框架语义学等理论。此外,本研究参考了詹卫东[2]的"广义配价模式"(Generalized Valency Mode)以及毕玉德[3]的"韩国语句法语义信息词典"的语义组织结构,词典中包括了形容词基本分类、形容词语义类、形容词情感分类、补足语语义角色分类、补足语语义特征描述、价数判定、句式及例句等语义

黄金柱(1980—),男,博士生,主要研究方向为自然语言处理、本体资源库建设,E-mail:hjz-johnsmith@126.com;李 峰(1982—),男,博士,工程师,主要研究方向为计算语言学、数据挖掘,E-mail:li-bopr@126.com;张克亮(1964—),男,教授,博士生导师,主要研究方向为自然语言处理、本体资源库建设等,E-mail:kliang99@sina.com。

信息。与基于配价的汉语语义词典(VCSD)不同的是,VDEA词典主要面向英语形容词的语义知识描述,针对性强且语义信息更加全面。

2.1 知识来源

VDEA词典主要面向自然语言处理,以英语形容词词条为中心,内容覆盖配价、语义类和语义格等词汇语义信息。词典主要参阅"A Valency Dictionary of English"(2004)[4]、《英语形容词搭配结构词典》(2004)、《英语形容词用法词典》(2005)、《形容词分类词典》(2004)、《褒义词词典》(2005)、《贬义词词典》(2005)、《汉语褒贬义词语用法词典》(2001)、金山词霸(2009 牛津版)、Lingoes 词典(2.6.1.0 版)以及 How-Net、WordNet 和北大汉语语言学研究中心公开语言资源[1)]。结合相关语义描述方法对所收集的词语资源进行抽取、加工和整理,共收集3170个英语形容词,对每个形容词进行详细的配价语义信息描述,在此基础上分别进行形式化表示,以满足人机两用的需要。

2.2 词典的设计

2.2.1 总体框架设计

本研究设计的配价词典框架与北大的基于配价的汉语语义词典(VCSD)有很大不同,例如在配价语义信息的基础上加入了面向情感分析的情感信息描述以及大量的句式和例句信息等。在数据库表的设定上整体可分为两大类,即主表和辅表。主表即形容词配价信息表,在参考现有文本及电子词典资源的基础上,结合汉语语义词典(VCSD)的构思进行设计,表内囊括英语形容词的配价语义信息以及相关情感信息,是词典的主要构成部分。创建辅表旨在方便形容词配价主表的形式化表示及应用,主要包括形容词语义类表、形容词情感类表、形容词基本分类表、补足语语义特征表、补足语语义角色表以及补足语形式类表等。

VDEA词典的总体结构如图1所示。

图1 配价词典的结构框架

2.2.2 表结构设计

在数据库整体结构确定后,要对配价信息主表以及辅表分别进行表结构设计,包括表的层次结构设定和属性字段的设定等,以下结合词典表结构图对配价信息主表以及相关辅表分别进行说明。

2.2.2.1 形容词配价信息主表

作为配价词典核心内容,形容词配价信息主表所包含的配价语义信息很丰富,因此对表的属性字段设定得比较精细,覆盖了词语、义项、所属次类、语义类、褒贬分类、情感类、释义、配价数、主体语义特征、客体语义特征、句式、句型结构以及大量例句等信息。表结构如图2所示。

ID	词语	所属次类	义项	释义	褒贬类	情感类	配价数	语义类	主体
31	abstemious	A[静态][A[+人][A[+自主]		节约的	+	F8	2	A32	[人]
32	abstract	A[静态][A[-人][A[+抽象]		抽象的	=	F8	1	A226	[内容]
33	abstruse	A[静态][A[-人][A[+抽象]		深奥的	=	F8	1	A226	[内容]
34	absurd	A[静态][A[-人][A[+抽象]		荒谬的	-	F82	1	A1	[事情]
35	abundant	A[静态][A[-人][A[+抽象]		大量的	=	F8	1	A2315	[物质]
36	abusive	A[动态][A[+人][A[-自主]		骂人的[攻击性的	-	F63	1	A32	[人,动物]
37	absurd	A[静态][A[-人][A[+抽象]		荒唐的	-	F63	1	A1	[事情]
38	abysmal	A[静态][A[-人][A[±抽象]		深不可测的	=	F8	1	A229	[思想[非生物]
39	academic	A[静态][A[-人][A[-抽象]		学术的[学院的	=	F8	1	A5	[领域]
40	accepted	A[静态][A[-人][A[-抽象]		公认的	=	F8	1	A225	[结果[事情]
41	accessible	A[静态][A[-人][A[-抽象]		容易达到的	=	F8	1	A229	[结果]
42	accidental	A[静态][A[-人][A[+抽象]		意外的	=	F8	1	A1	[问题[事情]
43	accomplished	A[动态][A[+人][A[-自主]		熟练的[精熟的	+	F21	1	A32	[人]
44	accountable	A[动态][A[+人][A[+自主]		负有责任的	-	F8	2	A32	[人]
45	accredited	A[静态][A[-人][A[+抽象]		可接受的[合格的	=	F8	1	A229	[结果[事情]
46	acephalous	A[静态][A[-人][A[+抽象]		无头的	=	F8	1	A34	[事情]
47	acidulous	A[动态][A[+人][A[+自主]		不快的	=	F3	1	A34	[人]
48	acknowledged	A[静态][A[-人][A[-抽象]		公认的	=	F8	1	A225	[结果[事情]
49	acquiescent	A[静态][A[-人][A[-抽象]		默许的	=	F8	1	A1	[结果[事情]
50	acquisitive	A[静态][A[-人][A[+抽象]		想获得的[可学到的	=	F8	1	A225	[属性]
51	acrid	A[静态][A[-人][A[-抽象]		辛辣的	=	F8	1	A224	[物质]
52	acting	A[静态][A[-人][A[+抽象]		代理的	=	F8	1	A225	[人]
53	acute	A[动态][A[+人][A[-自主]		敏大的[急性的	=	F71	1	A229	[程度[人]

图2 形容词配价信息表

根据配价语法以及格语法等语法理论,并参考相关知识库语义信息表示方法,每个字段所表示的语义信息如下。

(1)发音。

(2)所属次类:分别从动态与静态、述人与非述人、可控与非可控、抽象与非抽象、自主与非自主、感受与非感受6个方面对现有英语形容词进行二分,此层次分类属于"粗分类"[2,5]。

(3)义项。

(4)释义。

(5)褒贬类[6,12]:本研究把所有库内收集的英语形容词进行褒贬分类,主要分为褒义类、贬义类和中性类3种,分别用"+"、"-"和"="这3种符号进行表示。例如 able、abrupt 和 ablaze 这3个形容词划定为褒义类、贬义类和中性类,分别用"+"、"-"和"="表示。

(6)情感类[11]:对每个形容词进行情感类划分是机器进行情感分析的基础,为自然语言处理中的情感分析提供帮助。具体的情感划分如图3所示。

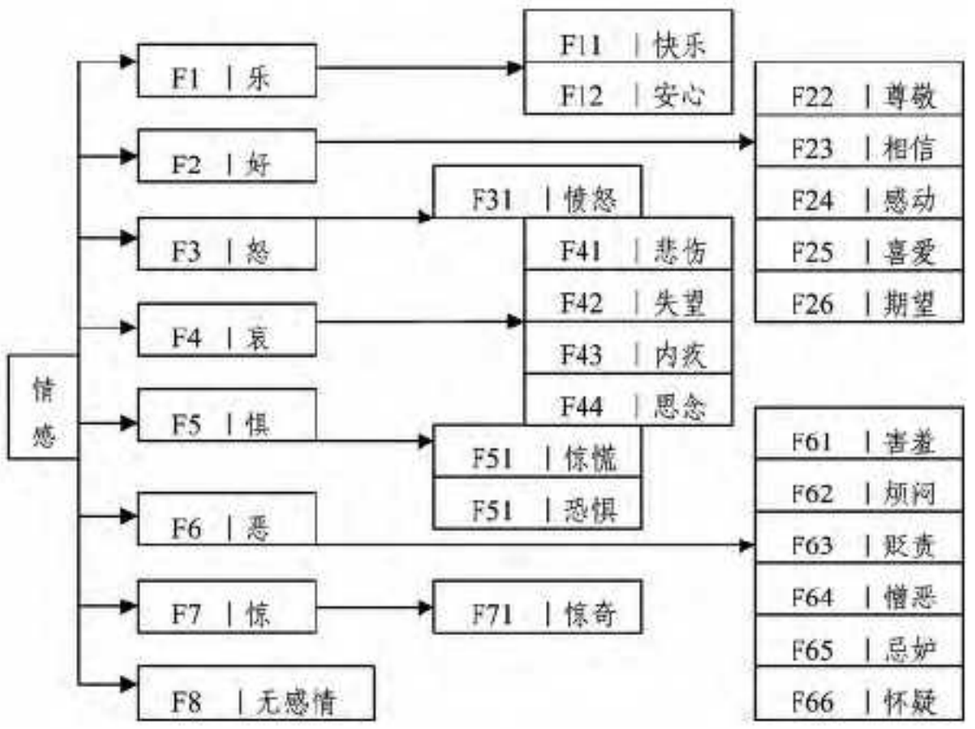

图3 形容词情感分类图

将不含情感色彩的形容词统一归入"非感情"类,分别用大写字母"F"与阿拉伯数字组合表示这8类形容词,例如在 abrupt、accomplished 和 amphibian 这3个形容词中,abrupt 表示的"鲁莽的"义项具有贬责义,用"F63"表示;accomplished 表赞扬义,用"F21"表示;而 amphibian 释义为"两栖动物的",不具有情感性,用"F8"表示。

(7)配价数[7,8]。

(8)语义类:形容词语义类又称作形容词的"语义特征",是对 VDEA 词典所有形容词进行的"细分类"。

(9)主体与客体:形容词的语义角色相对动词来说比较简

1) http://www.ccl.pku.edu.cn/ccl-sem-dict

单，主要从主体与客体两类分别进行补足语的语义特征描述。

(10)句式、句型结构与例句。VDEA 词典的句式与例句信息包括了“A Valency Dictionary of English”词典[1]的全部形容词配价信息，并在句式和例句基础上进行形式化表示。

2.2.2.2 VDEA 辅表

这里仅简要介绍两个重要辅表即形容词语义类表和形容词情感类表，其它相关辅表不再赘述。形容词语义类表是对形容词进行语义类细划分的基础。该表同样以“ID”编号为主键与主表建立联系，借助此表可以对主表形容词进行“细分类”并加以形式化表示。主表由形容词次类表、形容词次类 1 表和形容词次类 2 表 3 个表构成。形容词语义类表采用 Access 数据库的“子表嵌套”功能，由粗到细进行多层“子表嵌套”，将形容词的 5 大类及 35 小类清楚表示出来。表结构如图 4 所示。

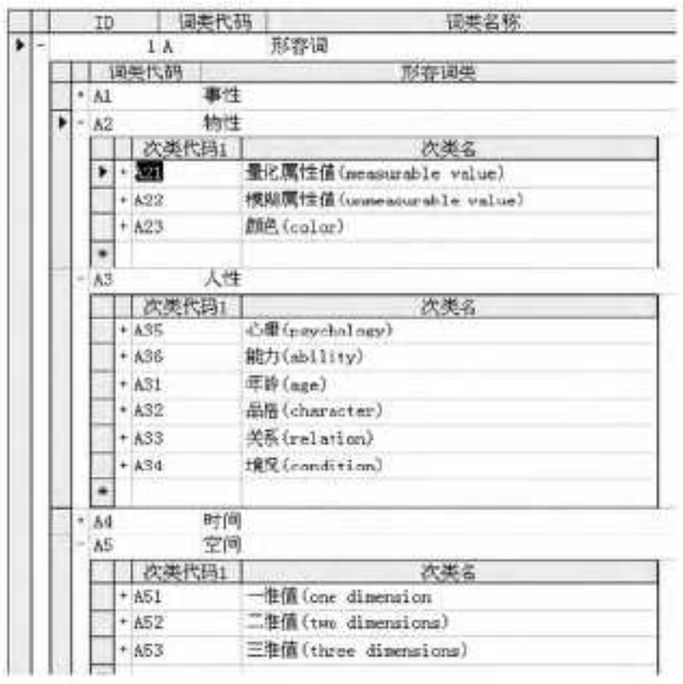

图 4 形容词语义类表

此外，形容词情感类表是 VDEA 词典应用于情感倾向性分析领域的重要辅表，其中包含了 7 个情感大类和 29 个情感小类，由“编号”、“代码”以及“情感类名”这 3 个属性字段组成，以“编号”作为表的主键与配价信息主表建立联系，在“代码”字段用单词“feeling”的首字母“F”与阿拉伯数字组合对所有情感类别进行形式化表示，“情感类名”描述了所有情感的名称。表结构如图 5 所示。

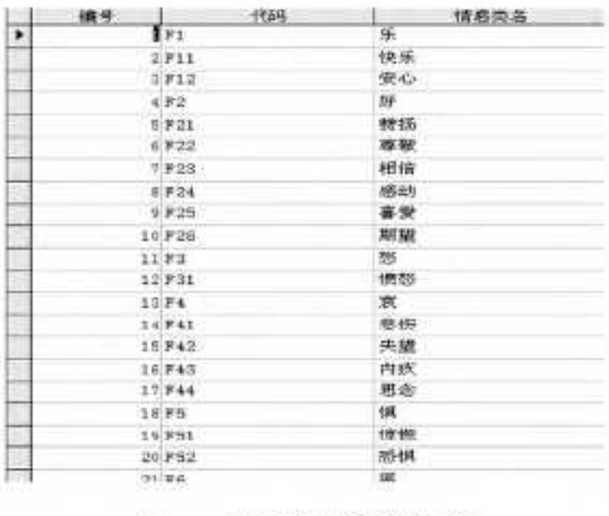

图 5 形容词情感类表

3 基于 VDEA 的词汇情感倾向分析

基于情感词典的情感倾向性分析方法很多，具体的分析方法因词典的特点不同而各异，例如周咏梅等、潘文彬、黄硕以及张成功等的研究。周咏梅等学者[9]基于自建中文情感词典(SLHS)进行词语正负情感倾向值计算；潘文彬[10]提出了对大规模情感词典进行切分的方法，从而降低了置信度较低的词语对情感计算的负面影响；黄硕[11]根据情感分类等方法进行情感知识库的构建，基于该知识库进行情感倾向计算；张成功[12]整理并构建了基础情感词典、领域词典、网络词典以及修饰词词典，并进行了短语、句子和篇章的极性计算分析。

VDEA 词典采用编码形式对英语形容词进行层次结构划分，从语义类、情感类、褒贬极性、配价关系以及语义特征等方面分别对英语形容词以及形容词补足语成分进行分类和语义信息描述，在 12 个基本分类基础上，将 3170 个英语形容词细分为 5 个大类、35 小类以及 30 个情感类，将补足语分为 139 个语义类和 25 个语义格，同时辅助知识库中收集形容词的同义、近义、反义以及短语信息，并分别对这些信息进行了极性分类。本文根据词典的组织结构得到形容词情感倾向性分析的相关理论基础：

(1)同一褒贬类和情感类的形容词具有相同或相近的情感倾向性，如 accomplished 和 adamant 这两个形容词都属于褒义词，情感上都属于“赞扬”类，用“F21”表示。

(2)同一语义类的形容词具有相近或相反极性，如 admirable 和 adverse 都属于“境况”语义类，用“A34”表示，但 admirable 为褒义，而 adverse 为贬义，极性相反。

(3)形容词同义词典包含的同义形容词褒贬极性相同，形容词反义词典则反之。

(4)形容词直接或间接影响形容词短语的褒贬极性。

本文提出的词汇情感倾向算法主要由以下步骤组成：

(1)特征词主观性选择。首先从 VDEA 词典中人工提取出具有明显褒贬倾向的形容词作为特征词，特征词的数量控制在合适的范围，因为基准词过多会导致情感特征不明显进而达不到“标杆”效果，而特征词过少又会导致“标杆”覆盖面不全的问题。

(2)特征词统计选择。利用 Google 等搜索引擎对人工选取的情感形容词进行“Hits”数统计，将“Hits”数作为形容词的词频权重对形容词进行二次筛选，对权重较低形容词给予删除处理，仅保留权重值较大的形容词，并最终得到 K 对褒贬特征词集。

(3)形容词语义距离计算。将双重挑选后的形容词作为特征词，将测试集形容词与 K 对褒贬特征词分别进行比较，根据两者在词典中的语义类编码求得语义距离。

(4)形容词情感距离计算。将双重挑选后的形容词作为特征词，将测试集形容词与 K 对褒贬特征词分别进行比较，根据两者在词典中的情感类编码求得情感距离。

(5)形容词相似度计算。结合特征形容词与测试形容词的语义距离和情感距离，求得两者的相似度值。

(6)形容词情感倾向性计算。设立合适阈值，对每个形容词的相似度计算结果进行统计，利用求差比较法得出最终结果并与阈值作比较，若大于阈值则归为褒义倾向，反之则归为贬义倾向。

基于 VDEA 词典的词语情感倾向值计算流程如图 6 所示。

1) 本研究利用扫描仪对 Herbst 的英语配价词典进行扫描，对扫描结果进行加工，从中提取出所有的英语形容词并进行整理，最后将获取的形容词配价信息直接导入 Access 数据库，从而构成了 VDEA 词典的语义知识基础。

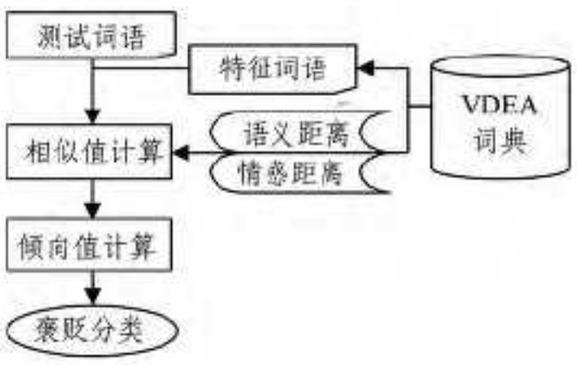

图 6　词语倾向性分析流程图

3.1　特征词的选取

VDEA 词典的多角度词语分类给特征词的划定带来了很大便利，特征词选取主要包括以下几步骤：

(1)首先以手工方式从词典中选定 600 个形容词。

(2)利用 Google 搜索引擎进行“Hits”数确定并加以比较，最终抽取“Hits”数相对较高的 200 个形容词。

(3)对初步选定的每个形容词按语义类、情感类以及褒贬极性 3 个层面进行挑选，对属于同一语义类、情感类的形容词进行部分删除处理，仅保留代表性较强且“Hits”数相对较高的形容词，最终从中抽取 20 对褒贬形容词作为特征词。

3.2　形容词语义相似度计算

刘群[13]在基于 HowNet 的词汇语义相似度计算方法研究中，把两个独立的词汇相似度问题归为两个概念间的相似问题[1)]，如对于 W_1 和 W_2 两个词，W_1 有 n 个义项且 W_2 有 m 个义项，则这两个词的相似度为所有义项间相似度的最大值，即：

$$Sim(W_1,W_2)=\mathrm{Max}\ Sim(S_{1i},S_{2j}),i\in[1,n],j\in[1,m] \tag{1}$$

而义项的相似度取决于义原的相似度，并且每个义项都由 4 部分组成：独立义原、其它义原、关系义原和符号义原，在进行义项的相似度计算时必须考虑这 4 个因素。义原相似度算法为：

$$Sim(P_1,P_2)=A/(D+A) \tag{2}$$

P_1 和 P_2 是义项的两个义原，A 是 P_1 和 P_2 在义原层面上的路径长度，D 为可调节参数，因此得出最终的两义项间的整体相似度公式为：

$$Sim(S_1,S_2)=\Sigma-(i-1)^4-\beta_i\Pi-(j-1)^iSim_j(S_1,S_2) \tag{3}$$

其中，$\beta_i(1\leqslant i\leqslant 4)$ 是可调节参数，并且满足条件：$\beta_1+\beta_2+\beta_3+\beta_4=1,\beta_1\geqslant\beta_2\geqslant\beta_3\geqslant\beta_4$。$Sim_1(S_1,S_2)$、$Sim_2(S_1,S_2)$、$Sim_3(S_1,S_2)$ 以及 $Sim_4(S_1,S_2)$ 分别表示义原的 4 个构件。

VDEA 词典在语义信息的组织结构上不同于 HowNet，并且主要针对形容词进行分析处理，对此，本文采取分而治之的策略：对于动词、名词及副词，完全以 HowNet 情感词典为基础资源；对于形容词，则基于两个词典的形容词情感知识。

基于两个词典的形容词情感分析主要根据形容间的词典编码位置求距离值，进而计算词汇间的词义相似度，如式(4)和式(5)所示：

$$Sim_1(W,P_{ij})=A/(D_{1j}+A)\ ,i\in[1,2],j\in[1,20] \tag{4}$$

$$Sim_2(W,P_{ij})=A/(D_{2j}+A),i\in[1,2],j\in[1,20] \tag{5}$$

其中，W 为测试形容词，P_{1j} 表示褒义特征形容词，P_{2j} 表示贬义特征形容词，A 是可调节参数，D_{1j}、D_{2j} 表示 W 与 P_{1j}、P_{2j} 的语义距离和情感距离，Sim_1 表示语义相似度，Sim_2 表示情感相似度。通过式(4)和式(5)得到测试形容词的相似度为两个序列值，这里取两个序列中的最大值作为测试形容词与特征词的相似度值，如式(6)和式(7)所示：

$$Sim_1(W,P_i)=\mathrm{Max}\ Sim_1(W,P_{ij}),i\in[1,2],j\in[1,20] \tag{6}$$

$$Sim_2(W,P_i)=\mathrm{Max}\ Sim_2(W,P_{ij})i\in[1,2],j\in[1,20] \tag{7}$$

$Sim_1(W,P_i)$ 表示词汇语义相似度最大值，$Sim_2(W,P_i)$ 表示词汇情感相似度最大值。在此基础上将语义相似度与情感相似度进行综合处理，得出测试形容词与特征词间的综合相似度，即：

$$Sim(W,P)=\alpha Sim_1(W,P_i)+\beta Sim_2(W,P_i),\alpha/\beta\in[0,1],\alpha<\beta\ \text{且}\ i\in[1,2] \tag{8}$$

其中，$Sim(W,P)$ 表示最终的综合相似度，α 和 β 为独立的可调节参数，在 0 与 1 间取值，且 α 和 β 的取值与形容词的特点有关，在判定形容词褒贬倾向性时，语义对词汇倾向的影响力要小于情感。

3.3　情感倾向性计算

在得到两个形容词间的综合相似度值的基础上，可以通过特征词的褒贬极性来推定测试词的褒贬倾向性。朱嫣岚[14]利用 HowNet 进行词汇的褒贬倾向研究，杨昱昺和吴贤伟[15]对朱嫣岚的词汇倾向性算法进行了改进，本研究结合 VDEA 词典的特点对杨昱昺的算法进行修改，引入情感相似最大值对倾向值进行调节，得出式(9)：

$$Orientation(W)=[\gamma S_1+\delta Sim_2(W,P_i)]-[\gamma S_2+\delta Sim_2(W,P_i)] \tag{9}$$

式中，S_1 表示测试形容词与 20 个褒义特征词的综合相似度之和，S_2 表示测试形容词与 20 个贬义特征词的综合相似度之和，γ 和 δ 是可调节参数，用来对综合相似度和的值以及情感相似最大值进行适当调整，以提高计算准确性。

基于知网的词汇情感计算是通过计算词汇在知网中的距离值来实现的，本文则采取同样的思想，通过词汇在 VDEA 词典中的位置来计算词汇间的语义距离和情感距离。稍有改进的是在式(9)中，使用情感相似最大值替换语义相似最大值，这样便于发挥 VDEA 在情感计算中的作用，从而提高情感计算效能和结果的准确性。

3.4　实验结果与分析

按本文特征词选取方法从 VDEA 词典中抽取出 20 对特征形容词，从本词典中抽取 5 对形容词作为测试词汇，利用式(9)进行倾向性计算，将阈值设定为 0，倾向值大于 0 的为褒义形容词，低于 0 的为贬义形容词。在特征词的选用上本文分为 3 组进行实验，分别以 5 对、10 对、20 对特征词进行计算。以 5 对特征词为例分析如下，特征形容词及编码表(语义|情感)如表 1 所列。

表 1　特征形容词及编码表(语义|情感)

	5 对特征形容词
褒	Able(A36|F23)　brave(A32|F21)　new(A225|F8)　adamant(A32|F21)　wary(A32|F8)
贬	Mad(A34|F63)　terrible(A1|F71)　cruel(A32|F63)　old(A31|F8)　poor(A34|F8)

以测试形容词“Good(A228|F2)”为例，结合表 1 特征词，

1) http://hownet.kooge.com/hownet/smartfinder/index.jsp?hsign=6

根据词语在 VDEA 词典中所处位置可以得到与特征词的语义距离和情感距离，如表 2 所列。

表 2 语义情感距离表

语义距离	褒义	17,13,3,13,13
	贬义	15,12,13,12,15
情感距离	褒义	3,1,7,3,3
	贬义	7,8,7,6,6

结合词汇情感倾向计算分式进行情感倾向计算，经对比分析，在式(3)中当参数值"$A \geqslant 4$"时得出的相似度更趋合理，在此将 A 值定为 4，α 和 β 在[0.5,1]之间取值时结果趋于准确，经对比分析分别定为 0.6 和 0.8，γ 和 δ 是对综合倾向值与情感倾向最大值之和进行调节的参数，经比较将参数 γ 和 δ 值分别定为 0.5。最终得出 good 的倾向值为 0.39，大于阈值 0，为褒义形容词。同理对其它测试形容词与特征形容词作以上距离计算，并进行倾向值计算，实验结果如表 3 所列。

表 3 不同特征词集褒贬倾向性准确率(%)

5 对特征形容词			10 对特征形容词			20 对特征形容词		
褒义词准确	贬义词准确率	平均准确率	褒义词准确	贬义词准确率	平均准确率	褒义词准确	贬义词准确率	平均准确率
80	87	84.5	83	92	87.5	90.5	98	94.25

因为所选特征词数量以及测试词数量较少，所以整体效果一般，得出的准确率偏低。从实验结果来看，褒义形容词准确率比贬义形容词准确率低，这与褒贬形容词的特征有关。提高准确度的方法有两点，首先要提高特征形容词的数量以及覆盖面，其次是在算法上进行优化，尤其是在可调节参数的设定上要进行多次测试并将结果进行分析比较，使参数值更趋合理，从而提高词汇情感倾向性的准确率。

结束语　本文以英语形容词语义配价和情感信息描述为主要研究对象，为英语形容词语义信息和情感信息的分析处理探索新途径，在一定程度上填补了形容词配价研究的空白，拓宽了配价语法的研究范围。此外，基于本词典提出了词汇情感倾向性分析模型，在词汇情感倾向分析中引进情感相似最大值作为调节参数，提出了将语义距离和情感距离相结合进行词汇情感倾分析的方法。面向情感倾向分析的应用是构建本词典的重要目标，在不断丰富和完善配价词典的同时，还需积极探索词典在其它具体项目中的应用，如服务于政治、军事领域的语义检索、信息抽取和机器翻译等。

参 考 文 献

[1] Huang Jin-zhu, Yi Mian-zhu. The Design and Implementation of VDEA[C]// 4th International Universal Communication Symposium Proceedings. Beijing, 2010: 324-330

[2] 詹卫东. 一个汉语语义知识表达框架：广义配价模式[OL]. [2014-10-21]. http://ccl.pku.edu.cn/doubtfire

[3] 毕玉德. 面向韩文信息处理的谓词句法语义信息词典的构建[J]. 解放军外语学院学报, 2004, 25(4): 53-57

[4] Herbst T, et al. A Valency Dictionary of English [J]. International Journal of Lexicography, 2009, 22(1): 55-85

[5] 傅玉芳. 常用形容词分类词典[D]. 上海：上海大学出版社, 2004

[6] Bkyt T, et al. Polarity classification of celebrity coverage in the Chinese press[EB/OL]. [2014-08-12]. http://analysis.mitre.org/Proceedings/index.html

[7] 王付君. 性质形容词的价[J]. 长春教育学院学报, 2012, 28(1): 42-44

[8] 邱天. 现代汉语双价形容词研究[D]. 长春：东北师范大学, 2006

[9] 周咏梅，杨佳能. 面向文本情感分析的中文情感词典构建方法[J]. 山东大学学报, 2013, 23(6): 27-33

[10] 潘文彬. 基于情感词典的中文句子情感倾向分析[D]. 北京：北京邮电大学, 2011

[11] 黄硕. 中文情感知识库的构建与应用[D]. 北京：北京邮电大学, 2013

[12] 张成功，刘培玉，等. 一种基于极性词典的情感分析方法[J]. 山东大学学报, 2012, 47(3): 47-50

[13] 刘群，李素建. 基于《知网》的词汇语义相似度的计算[OL]. [2015-03-28]. http://www.wenku.baidu.com/view/6213af995/e79b8968022660.html

[14] 朱嫣岚，闵锦，等. 基于 HowNet 的词汇语义倾向计算[J]. 中文信息学报, 2007, 20(1): 14-20

[15] 杨昱昺，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

(上接第 417 页)

[9] Chen J, Hsu W, Lee M L, et al. Increasing Confidence of Protein Interactomes Using Network Topological metrics[J]. Bioinformaties, 2006, 22: 1998-2004

[10] Brun C, Chevenet F, Martin D, et al. Functional Classification of Proteins for the Prediction of Cellular Function from a Protein-Protein Interaction Network[J]. Genome Biol, 2003, 5(1): 1-13

[11] Chua H N, Sung W K, Wong L. Exploiting Indirect Neighbors and Topological Weight to Predict Protein Function from Protein-Protein interactions[J]. Bioinformatics, 2006, 22: 1623-1630

[12] Chou K C. Prediction of Protein Cellular Attributes Using Pseudo-amino Acid Composition [J]. Proteins: Structure, Function, and Bioinformatics, 2001, 43(3): 246-255

[13] http://www.csie.ntu.edu.tw/~cjlin

[14] Zhu S, Yu K, Chi Y, et al. Combining Content and Link for Classification Using Matrix Factorization[C]// Proceedings of SIGIR'07. Amsterdam, The Netherlands, 2007: 487-494

[15] Cho R J, Campbell M J, et al. A Genome-Wide Transcriptional Analysis of the Mitotic Cell Cycle[J]. Molecular Cell, 1998, 2(1): 65-73

[16] Mewes H W, Frishman D, et al. MIPS: a database for genomes and protein sequences[J]. Nucleic Acids Res, 2002, 30(1): 31-34

[17] Backstrom L, Leskovec J. Supervised Random Walks: Predicting and Recommending Links in Social Networks[C]// Proceedings of the fourth ACM International Conference on Web Search and Data Mining, 2011. Hong Kong, 2011: 635-644

[18] Hanley J A, McNeil B J. The meaning and use of the area under a receiver operating characteristic(ROC) curve[J]. Radiology, 1982, 143: 29-36