

基于认知语言学的自然语言语义表示方法

叶锡君 尹岩

(南京农业大学信息科技学院 南京 210095)

摘要 语义网对语义理解和常识推理是有效的,但粗粒度语义无法表示复杂的对象间的关系。认知语言学提供了表示复杂对象关系的方法,但由于其抽象性而难以应用于自然语言处理。文中提出一种基于认知语言学理论的自然语言语义表示方法。该方法使用意象图式表示语义网中对象节点间的关系,使用属性空间表示可以数值化的语义(如颜色)。在此基础上,将语义网和属性空间的构建和修改过程转化为语义操作序列,实现了句子语义的动态重构。相比于传统的基于语义网的语义表示方法,提出的方法能够表示动态关系,具有更强的推理能力。文中通过实例证明了这一方法的可行性。

关键词 认知语言学,语义网,属性空间,意象图式,语义操作,推理

中图分类号 TP391.1 文献标识码 A

Natural Language Semantic Representation Based on Cognitive Linguistics

YE Xi-jun YIN Yan

(School of Information Sciences and Technology, Nanjing Agricultural University, Nanjing 210095, China)

Abstract Semantic network is valid to semantic understanding and common sense reasoning, but coarse-grained semantic is difficult to represent complex relationship between objects. Though cognitive linguistics provides approaches to represent complex object-relational, it's difficult to support natural language processing because of its abstract. This paper presented a method of natural language semantic representation based on cognitive linguistics. The method uses image schema to represent the relationship between object nodes in semantic network, and attribute spaces to represent the semantic of object which can be digitized (such as color). On this basis, we transformed the building and modifying process on semantic network and attribute space into a sequence of semantic operations, so that we could reconstruct the semantics of sentences dynamically. Compared to the traditional method of semantic representation based on semantic network, the proposed method can express dynamic relationship and has stronger reasoning ability. The paper gave examples to prove the feasibility of this approach.

Keywords Cognitive linguistics, Semantic network, Attribute space, Image schema, Semantic operation, Reasoning

1 引言

在过去的近二十年中,基于经验主义的方法在自然语言处理领域获得了巨大的成功,但其成果主要集中于简单的任务,对于诸如语义理解这样的复杂任务较为乏力。考虑到智力正常的人可以轻松地完成各种复杂的语言任务,这类任务的解决有必要借鉴认知科学,尤其是认知语言学的各种研究成果,对语言机制内在特性的探究也应当超过经典语言学、计算语言学的范畴^[1]。

由于认知语言学理论通常比较抽象,将认知语言学的研究成果应用于自然语言处理的研究还很少。国内方面,袁毓林进行了不少有益的探索^[2,3]。他在文献^[3]中讨论了不同容器隐喻的心理表征问题,在此基础上使用一阶谓词逻辑给出了“满”和“全”的语义规则,这些规则经过调整后就可以作为自然语言处理系统所使用的规则。陈振宇^[4,5]从认知科学

的角度研究了汉语的时间系统和疑问系统,同样给出了可以应用于自然语言处理的运算规则。

语义网是一种基于图结构的知识表示方法,已经证明这种方法对语义理解和常识推理非常有效。但语义网也有一些明显的缺陷,语义网中节点间关系所表示的语义粒度过大,导致语义网无法表示复杂的对象间关系,并且限制了语义网的推理能力;对于一些语义,单纯使用图结构无法有效表示语义间的接近程度^[6]。一些已有的方法可以在一定程度上克服语义网的这些缺陷。对于前者,有很多语义网研究者致力于将语义网节点间的关系标准化,获得独立于语言个体或短语表达的特质,以提升使用语言结构进行推理的能力^[7]。对于后者,朱海平^[8]利用基于数值空间的模糊语义匹配实现了颜色概念的匹配,拓展了语义网的语义表示能力。

本文描述了一种基于认知语言学意象图式理论的语义网(简称语义图)。意象图式是认知语言学的主要研究对象之

本文受国家自然科学基金(31301691)资助。

叶锡君 博士,副教授,主要研究领域为生物信息学、数据挖掘及知识发现、计算机网络技术, E-mail: yexj@njau.edu.cn; 尹岩 硕士生,主要研究领域为人工智能、自然语言理解。

一,它本身具有意义,它的结构具有逻辑性,是最重要的语义结构^[9]。语义图使用意象图式提供的粒度更小的语义作为原子语义,用这个原子语义集合以统一的方式表示语义网中各节点间的关系,使用属性空间表示各种不可再分的原子对象的语义,并最终将两者统一在基于意象图式的语义操作之下。相比于传统的语义网,我们提出的语义图具有更灵活的语义扩展能力、更强的语义表示能力和推理能力。

2 语义要素

描述客观世界的自然语言的语义应该包括 3 种要素:对象、属性以及对象间的关系^[10,11]。这些要素都是语义单位,对应于语义图的节点和属性空间中的元素。语义图是一种使用关系节点表示对象间关系的语义网。属性空间是在一次语义理解过程中出现的可以借助数值空间表示的所有对象的集合。

三要素的定义如下:

定义 1 对象是可被独立认知的客观存在。

定义 2 一个对象可由若干个的其他对象构成,作为构成元素的对象是作为主体的对象的属性。

定义 3 对象间的关系是可用基本意象图式表示的对象和对象的联系或状态。

充当属性的对象本身具有语义,但其语义通常必须通过其依附对象才能被认知。感知觉属性源于人类的感知觉信息,是不可再分的原子对象,是原子语义。除感知觉属性外,原子对象还包括:人类在进化过程中发展起来的心理适配器^[12](文中也称“心理属性”),如“美”、“恐惧”、“欢乐”;数量属性,如“1”、“2”。属于基本层概念^[13,14]的对象(指那些在日常生活中我们能够直接看到、听到、摸到的对象)通常由感知觉属性直接构成,而复杂对象则由较简单的对象构成。

对象间的关系由意象图式提供的语义表示。简单的对象间关系可由基本意象图式表示。由基本意象图式复合而成的复杂意象图式可以表示复杂的对象间关系。

3 基于数值空间的对象语义的表示及计算

3.1 基于数值空间的对象语义的表示

能够用数值表示其语义的对象都是上文提到的原子对象。以颜色为例,可以使用基于 RGB 空间的三元组 $\langle R, G, B \rangle$ 表示颜色特征的语义,三元组的每个分量都是整数,取值范围为 $[0, 255]$ 。如“红色(红)”可以表示为 $\langle 255, 0, 0 \rangle$ 。一些原子对象与颜色类似,表示其语义的数值空间超过一维,如味觉、空间位置。另外一些原子对象的语义可以用单一维度表示,如长度、温度和质量等。还有少数原子对象的语义在直觉上难以转化为数值,如基本形状。这类对象可以通过选用特定特征将语义转化为数值空间中的点。对于形状,可以使用形状的边数或是组成这一形状的三角形的个数 n 表示其语义。

基于数值空间的对象语义表示方法有如下特点:

1. 使用数值空间表示语义,但不要求具有同样的形式。这里的形式包括维数和取值范围。

2. 同一数值空间中不同对象语义之间的关系可以通过计算得到。

必须指出的是,有些基于经典范畴理论的语义表示方法

同样使用数值表示语义,比如性别可分为“男”和“女”,分别用 1 和 0 表示。但在认知语言学中,性别概念涉及复杂的心理模型,并不属于原子对象。

3.2 基于数值空间的对象语义的计算

同一属性空间内的对象可以用欧式距离直观地表示它们语义的差别。同一属性空间内的两个对象语义距离的计算公式为:

$$d_x = \sqrt{\sum_{i=1}^n (a_i - b_i)^2} \quad (1)$$

a_i 和 b_i 表示两个对象 a 和 b 在维度 i 上的值。 n 为属性空间的维数。

同一属性空间内的两个对象在同一维度上的语义距离的计算公式为:

$$d_x = a_i - b_i \quad (2)$$

a_i 和 b_i 表示两个对象的语义 a 和 b 在维度 i 上的值。

$$\text{compare-one}(a, b) = \begin{cases} -1, & d_x < 0 \\ 0, & d_x = 0 \\ 1, & d_x > 0 \end{cases} \quad (3)$$

d_{\max} 表示对象语义在这一维度上取值范围的上界。

若已知对象 a 在维度 i 的语义 a_i , 对象 a 和对象 b 在维度 i 语义的差值 Δd , 那么对象 b 在维度 i 的语义为:

$$b_i = \begin{cases} d_{\max}, & a_i + \Delta d > d_{\max} \\ a_i + \Delta d, & 0 < a_i + \Delta d < d_{\max} \\ 0, & a_i + \Delta d \leq 0 \end{cases} \quad (4)$$

除了可以用 n 元组表示外,对象的语义还可以转化为属性空间内表示其语义的点与原点的距离:

$$V_+ = \sqrt{\sum_{i=1}^n a_i^2} \quad (5)$$

V_{\max} 表示这某一属性空间内所能表示的最大语义值。

设对象 a 和对象 b 的语义分别为 v_{a+} 和 v_{b+} , 则有:

$$\text{compare-all}(a, b) = \begin{cases} -1, & v_{a+} < v_{b+} \\ 0, & v_{a+} = v_{b+} \\ 1, & v_{a+} > v_{b+} \end{cases} \quad (6)$$

若已知对象 a 的语义 v_a 以及对象 a 和对象 b 的语义的差值 Δv , 那么对象 b 的语义为:

$$v_b = \begin{cases} v_{\max}, & v_a + \Delta v \geq v_{\max} \\ v_a + \Delta v, & 0 < v_a + \Delta v < v_{\max} \\ 0, & v_a + \Delta v \leq 0 \end{cases} \quad (7)$$

在推理过程中,式(4)中的 Δd 和式(7)中的 Δv 根据属性空间的不同采用不同的默认值,通常可以取 $d_{\max}/10$ 及 $v_{\max}/10$ 。

3.3 实例

我们用 RGB 表示颜色词汇的语义。在 RGB 空间中, $\langle 0, 0, 0 \rangle$ 表示“黑”而 $\langle 255, 255, 255 \rangle$ 表示“白”,所以越“浅”的颜色对应的语义值反而越大。设“红/红色”的语义为 $\langle 255, 0, 0 \rangle$, 根据上一节的公式,可以计算出“浅红”的语义为 $\langle 255, 26, 26 \rangle$ 。

4 基于意象图式的对象和对象间关系的表示及推理

4.1 对象与对象间的关系的表示

语义图使用 3 种节点:对象节点、关系节点、力量_动态节点。对象节点表示对象,关系节点和力量_动态节点表示对象

间的关系。节点之间用有向弧连接。同概念图^[15]类似,语义图的弧不带标签。弧的方向决定了对象和其他对象在特定关系中的角色。关系节点的语义来自于基本意象图式。Johnson和Lakoff总结了部分具有代表性的基本意象图式,表1是对这些意象图式的归纳^[9]。表中部分的意象图式的语义难以直接利用,本文对这类意象图式进行了修改和扩展。

表1 基本意象图式的语义

空间	上下	力量_动态	阻碍
	前后		转移
	左右		吸引
	远近		比较
	中心_边缘		匹配
等级	接触	辨认	移动
	路径		循环
	容纳		过程
容器	内外	存在	目标
	表面		封闭空间
	空_满		物质_可数
	内容		部分_整体
力量_动态	平衡	整体/多样	连接
	对抗		重复
	强迫		合并
	制止		集合
	成为可能		分裂

1. 对于类似“匹配”这类表示某种操作的意象图式进行了修改,将操作的“结果”作为关系节点的语义,对于“匹配”,有“相同/不相同”,对于“比较”则有“大于/相等/小于”。

2. 对某些意象图式进行了基于隐喻的扩展,使之更容易理解。比如在容器图式类型中添加“拥有”意象图式。

图1是“我希望见到她”的语义图,已省略了不必要的关系节点和空对象节点。

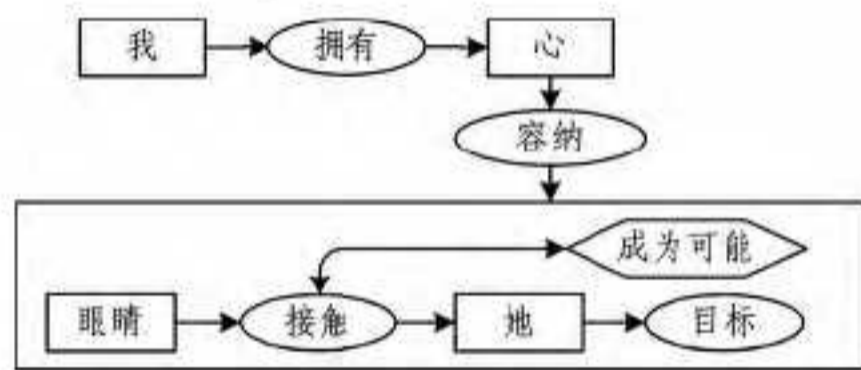


图1 “我希望见到她”的语义图

在利用意象图式时,还发现了它的下述性质:

1. 对象本身能够看作为一个容器图式,所以对象能够包含其他对象。定义2体现了这一性质。借助这种特性,语义图得以支持自然语言的递归性质。

2. 一些关系所涉及的对象是默认的,往往需要依据常识获得。比如“跑”有“位移”的意思,涉及的对象除了运动的主体外还包括作为参照物的地面。

3. 很多意象图式(如空间意象图式、力量_动态意象图式等)的语义需要结合各类属性空间才能完整表示。

4. 有些意象图式的语义必须依靠时间特征才能完整表示,比如“重复”、“循环”。另外一些意象图式难以仅用语义图表示,比如“封闭空间”。这些意象图式也十分重要,但使用它们则超出语义图目前的能力。

4.2 基于力量_动态意象图式的推理

力量_动态意象图式较为特殊。Talmy^[16]于1988年重点论述了这类意象图式。他认为,力对物体产生影响(如移动、克服阻力、越过障碍等)形成的意象图式,在人类的认知和语

言形成过程中起着核心的和普遍的作用。力有多种分类,不但有表示物理意义的力,还有社会和心理方面的力。在语义图中,动态_力量节点是语义图引入推理规则的标志,它连接了与规则相关的各个关系节点。推理规则有两方面的作用:根据规则添加动态_力量关系中涉及的对象节点和关系节点,作为属性空间内对象语义计算的依据(类似于3.2节给出的公式)。

以最简单的基于稳定状态(steady-state)力量_动态意象图式的推理为例,这类意象图式包括了4个要素:主角(Agonist)、反角(Antagonist)、主角和反角之间的力、主角和反角各自的运动状态。主角和反角默认是相互接触的,当主角和反角分开后,施力受力的关系会消失,但运动状态会保留下来。主角和反角是意象图式所表示的关系中的对象,与语法的主语宾语之间不存在固定的对应关系,甚至仅会作为默认对象出现。

基于稳定状态力量_动态意象图式的推理规则可以用自然语言描述为:

1. 如果主角和反角相互接触,且主角对反角的力大于反角对主角的力,则反角的运动状态改变。
2. 如果主角和反角相互接触,且主角对反角的力小于或等于反角对主角的力,则反角的运动状态不变。
3. 如果主角和反角不接触,则反角的运动状态不变。

力量_动态意象图式支持反向推理,其规则可用自然语言描述为:

1. 如果反角的运动状态改变,则主角和反角相互接触,且主角对反角的力大于反角对主角的力。
2. 如果反角的运动状态不改变,则主角和反角不接触,或主角对反角的力大于反角对主角的力。

4.3 实例

若不考虑力量_动态意象图式,“小明推箱子”的语义可以用图2表示。加入动态_力量节点后,变为图3所示。

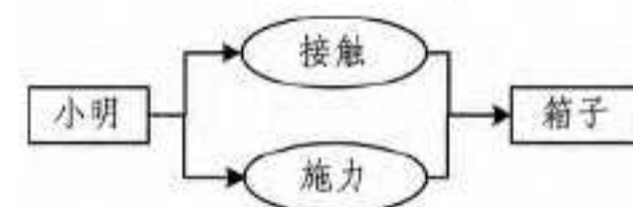


图2 “小明推箱子”未加入动态_力量节点

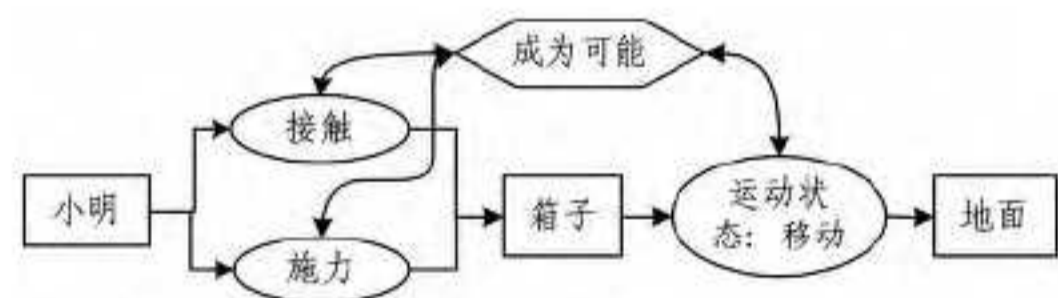


图3 “小明推箱子”加入动态_力量节点

图3中的六边形节点“成为可能”就是力量_动态节点,它与其他3个关系节点相互连接。图中还包括了另外两个图2中没有的节点:关系节点“运动状态”和对象节点“地面”。“地面”对象节点就是“移动”关系中的默认对象,作为“箱子”运动状态的参照物。而关系节点“运动状态”则表示了“箱子”可能具有的状态,在本例中包括“移动”和“不移动(静止)”两种,默认是“移动”。

基于力量_动态意象图式的推理规则要求在原本语义图和属性空间上添加规则中涉及的其他对象节点、关系节点以及必要的链接。在例子中就是“运动状态”节点和“地面”节点。类似地,对于句子“箱子移动”,语义图中存在“箱子”,“移

动”和“地面”₃个节点,还要根据规则添加默认的“施力者”节点,以及关系节点“接触”和“施力”。

基于力量_动态意象图式的推理过程表现为,若读入的下一个句子是“箱子静止”,可以根据规则修改力量属性空间的值,以表示小明对箱子的力的值小于或等于箱子对小明的力的值。下文详细介绍推理过程的实现方法。

5 语义图和属性空间的动态构建和修改

5.1 原子语义操作

言语是自然语言的具体实现,是词汇的线性序列,词汇具有语义,人类理解言语语义的过程可以看作顺序地处理词汇语义的过程。本文使用语义图和属性空间表示语义,所以言语语义的表示和理解过程应当转化为对语义图和属性空间的动态构建和修改。为了实现这种转化,我们把词汇作为针对语义图和属性空间的特定操作的标志。如表示对象的词汇“兔子”对应的操作是:1.在语义图上建立一个对象节点;2.填充对象节点所关联的符号为“兔子”。表示对象间关系的词汇“下/下边/下面”对应的操作是:在空间位置空间中(一个用以表示空间位置的属性空间),根据已经确定位置的对象A的位置计算尚未确定位置的对象B的位置,并将计算结果作为B的位置的值。

针对语义图或属性空间的不可再分的操作称为原子语义操作。这些操作包括在语义图上添加节点,在语义图上的不同节点间建立连接,修改属性空间中某个对象的语义值等。这里给出针对语义图的部分原子语义操作,见表2。

表2 语义图上的原子语义操作

操作名	描述
ActiveObjectNode	在语义图上添加一个空对象节点
AssignmentObject	对没有定义的对象节点填充语义(符号)
ActiveRelationNode	在语义图上添加一个空的关系节点
AssignmentRelation	对没有定义的关系节点填充语义(符号)
LinkObjectAndRelation	将对象节点与关系节点连接

根据3.2节给出的公式,可以得到针对属性空间的部分原子操作,见表3。

表3 属性空间上的原子语义操作

操作	描述
Compare	比较两个对象的语义,根据结果返回值
CompareInOneDemension	比较两个特定维度上的语义,根据结果返回值
Change	给对象的语义加上一个特定的值
ChangeInOneDemension	给对象特定维度上的语义加上一个值
ActiveObject	建立一个对象的语义,并将其设定为默认值

除了上述类型的原子语义操作外,在具体的计算机程序中,还需要一些起辅助作用的原子语义操作,比如取得对象的语义值(GetSemanticValue),设定对象的语义值(SetSemanticValue)。

5.2 语义操作序列

由若干个原子语义操作组成的有序序列称为语义操作序列。实现一个特定的语义操作序列,就可以实现对应语义图和属性空间的动态构建和修改,进而实现对应词汇或句子的语义的表示和理解。由于意象图式表示对象间的关系,提供了关系节点的语义,因此意象图式同样可以转化为对应的语义操作序列。比如“接触”意象图式可以转化为表4所列的语义操作序列。

表4 “接触”的语义操作序列

序号	操作
1	ActiveRelationNode()
2	ActiveObjectNode()
3	AssignmentRelation(“接触”)
4	LinkObjectToRelation()
5	GetSemanticValue()
6	Change(“接触”)
7	SetSemanticValue()

表示关系的词汇的语义可以由意象图式表示,如“推”。其语义由“接触”和“施力”组成(为了简化说明,这里不考虑动态_力量意象图式)。存在这样的层次关系,意象图式的语义操作序列由原子语义操作构成,词汇的语义操作序列由意象图式的语义操作序列构成,句子的语义操作序列由词汇的语义操作序列构成。

构建或修改语义图和属性空间的过程还需要确定节点间链接的方向或对象在特定关系中扮演的角色。在汉语中,表示对象的词汇在句子中出现的顺序决定了这个对象在关系中的角色,这里引入激活强度来描述这种现象。激活强度是一个整数,取值范围是[1,7](根据文献[17],人类工作记忆的容量是 7 ± 2 ,这里取7)。在语义图上建立一个新的对象节点或找到一个已经存在的对象节点时,节点会被赋予最高激活强度,之后每读入一个词,其激活强度减1,直至1。关系节点通常仅与当前激活强度最高和次高的对象节点相连接,并且可以根据对象节点激活强度的不同确定链接的方向。表2给出的LinkObjectAndRelation可以分解为LinkObjectToRelation和LinkRelationToObject两个操作。

5.3 实例

仍旧以“小明推箱子”作为例子。为了简化问题,不考虑动态_力量节点的使用。

在表5中存在4个ActiveObjectNode操作。第一个和最后一个来自“存在”意象图式,用以在语义图中加入对象节点。中间两个分别来自“接触”和“施力”意象图式的语义操作序列。ActiveObjectNode操作在建立节点前会察看当前语义图中是否有已经有空的对象节点,若存在,则不会建立新的对象节点。

表5 构建图2的语义操作序列

序号	词汇	原子语义操作
1		ActiveObjectNode()
2	小明	AssignmentObject(“小明”)
3		ActiveRelationNode()
4		ActiveObjectNode()
5		AssignmentRelation(“接触”)
6		LinkObjectToRelation()
7		LinkRelationToObject()
8		GetSemanticValue()
9	推	Change(“接触”)
10		SetSemanticValue()
11		ActiveRelationNode()
12		ActiveObjectNode()
13		AssignmentRelation(“施力”)
14		LinkObjectToRelation()
15		LinkRelationToObject()
16	箱子	ActiveObjectNode()
17		AssignmentObject(“箱子”)

结束语 传统的语义网研究着眼于具有一定内部结构的静态对象的表示方法,我们的研究更注重词汇与语义的对应以及语义的动态生成。通过将意象图式提供的语义作为关系节点的语义,并加入属性空间表示各种基本属性,语义图具备了比传统语义网更强的语义表示能力。而将语义的表示和理解转化为特定语义操作序列的实现,则使语义图拥有更加灵活的推理能力和更强的语义扩展能力。

后续的研究包括如何处理一词多义现象,如何将隐喻加入到系统中以表示更丰富的抽象概念,以及如何恰当地表示两个对象之间的类属关系等。

参 考 文 献

[1] 危辉. 人工智能形式概念系统[M]. 北京: 科学出版社, 2011

[2] 袁毓林. 基于统计的语言处理模型的局限性[J]. 语言文字应用, 2004, 5(2): 99-107

[3] 袁毓林, 陈振宇, 张秀松, 等. 从认知假设到计算分析和程序实现——一种认知语言学研究的计算范式与技术路线[J]. 当代语言学, 2010, 12(2): 97-114

[4] 陈振宇. 时间系统的认知模型与运算[M]. 上海: 学林出版社, 2007

[5] 陈振宇. 疑问系统的认知模型与运算[M]. 上海: 学林出版社, 2010

[6] Kedad Z, Métais E. Ontology-Based Data Cleaning [M]//

Andersson B, Bergholtz M, Johannesson P. Natural Language Processing and Information Systems. Springer Berlin Heidelberg, 2002: 137-149

[7] Luger G F. 人工智能: 复杂问题求解的结构和策略[M]. 史忠植, 张银奎, 赵志崑, 等译. 北京: 机械工业出版社, 2006

[8] 朱海平. 基于概念图匹配的语义搜索[D]. 上海: 上海交通大学, 2006

[9] 王寅. 认知语言学[M]. 上海: 上海外语教育出版社, 2007

[10] Barwise J, Perry J. Semantic innocence and uncompromising situations[M]// Martinich A P, ed. The Philosophy of Language (2nd Edition). New York: Oxford University Press, 1990: 392-404

[11] Barwise J, Perry J. Situations and Attitudes[M]. Stanford, CA: CSLI Publications, 1999

[12] 巴斯 D M. 进化心理学: 心理的新科学(第二版)[M]. 熊哲宏, 张勇, 晏倩, 译. 华东师范大学出版社, 2007

[13] 朱跃. 语义论[M]. 北京: 北京大学出版社, 2006

[14] 张维鼎. 意义与认知范畴化[M]. 成都: 四川大学出版社, 2007

[15] Sowa J F. Conceptual structures: Information processing in mind and machine[M]. Addison-Wesley, 1984

[16] Talmy L. Force dynamics in language and cognition [J]. Cognitive Science, 1988, 12(1): 49-100

[17] Sternberg R J. 认知心理学(第三版)[M]. 杨炳钧, 陈燕, 邹枝玲, 译. 北京: 中国轻工业出版社, 2006

(上接第 90 页)

[5] 刘华. 基于关键短语的文本内容标引研究[D]. 北京: 北京语言大学, 2005

[6] 韩艳. 基于统计的中文文本关键短语自动抽取方法研究[D]. 苏州: 苏州大学, 2009

[7] Mihalcea R, Tarau P. TextRank: Bringing Order into Texts [C]// Proceedings of EMNLP. 2004: 404-411

[8] 刘知远. 基于文档主题结构的关键词抽取方法研究[D]. 北京: 清华大学, 2011

[9] 方俊, 郭雷, 王晓东. 基于语义的关键词提取算法[J]. 计算机科学, 2008, 35(6): 148-151

[10] 索红光, 刘玉树. 一种基于词汇链的关键词抽取方法[J]. 中文信息学报, 2006, 20(6): 25-30

[11] 胡燕, 吴虎子, 钟璐. 中文文本分类中基于词性的特征提取方法研究[J]. 武汉理工大学学报, 2007, 4

[12] 赵军, 黄吕宁. 汉语基本名词短语结构分析模型[J]. 计算机学报, 1999, 22(2): 141-146

[13] 赵蕾蕾. 基于词和基本短语模式的特征提取方法[D]. 保定: 河北大学, 2009

[14] 王军. 词表的自动丰富——从元数据中提取关键词及其定位[J]. 中文信息学报, 2005, 19(6): 36-43

[15] Hulth A. Improved Automatic Keyword Extraction Given More Linguistic Knowledge[C]// Proceedings of EMNLP. 2003: 216-223

[16] Peter D. Turney, Learning Algorithms for Keyphrase Extraction

[J]. Information Retrieval, 2000, 2(4): 303-336

[17] Frank E, Paynter G W, Witten I H, et al. Domain-specific Keyphrase Extraction[C]// Proceedings of IJCAI. 1999: 668-673

[18] 李素建, 王厚峰, 俞士汶, 等. 关键词自动标引的最大熵模型应用研究[J]. 计算机学报, 2004, 27(9): 1192-1197

[19] Zhang K, Xu H, Tang J, et al. Keyword Extraction Using Support Vector Machine[C]// Proc. of the Seventh International Conference on Web-Age Information Management (WAIM 2006). 2006: 85-96

[20] Zhang Cheng-zhi, Wang Hui-lin, Liu Yao, et al. Automatic Keyword Extraction from Documents Using Conditional Random Fields[J]. Journal of Computational Information Systems, 2008, 4(3): 1169-1180

[21] 钱爱兵, 江岚. 基于改进 TFIDF 的中文网页关键词抽取——以新闻网页为例[J]. 情报理论与实践, 2008, 6

[22] 郑家恒, 卢娇丽. 关键词抽取方法的研究[J]. 计算机工程, 2005, 31(18)

[23] 都云程, 周伟, 韩艳锋, 等. 基于字同现频率的关键词自动抽取[J]. 北京信息科技大学学报, 2011, 26(6)

[24] 肖根胜. 改进 TFIDF 和谱分割的关键词自动抽取方法研究[D]. 武汉: 华中师范大学, 2012

[25] 赵鹏, 蔡庆生, 王清毅, 等. 一种基于复杂网络特征的中文文档关键词抽取算法[J]. 模式识别与人工智能, 2007, 20(6)

[26] 汪小帆, 李翔, 陈关荣. 复杂网络理论及其应用[M]. 北京: 清华大学出版社, 2006