

基于合作机制的汉语名词性隐喻理解方法

苏 畅, 周昌乐

(厦门大学人工智能研究所, 福建 厦门 361005)

摘 要: 从全新的合作角度来研究有效进行名词性隐喻理解的方法, 提出本体和喻体的三种合作方式。基于互动论的思想, 采用知识库与语料库相结合的方式来抽取隐喻句的表述义。实验结果表明, 提出的基于合作机制的隐喻表述义提取算法是有效的。

关键词: 合作; 隐喻理解; 自然语言理解

中图分类号: TP391 **文献标志码:** A **文章编号:** 1001-3695(2007)09-0067-03

Chinese nominal metaphor comprehension method based on cooperation mechanism

SU Chang, ZHOU Chang-le

(Institute of Artificial Intelligence, Xiamen University, Xiamen Fujian 361005, China)

Abstract: The paper put forward a novel method of nominal metaphor comprehension which was based on the mechanism of cooperation. It proposed three manners of cooperation between the tenor and the vehicle. The comprehension system was based on the idea of interaction. The result shows that the method based on the mechanism of cooperation is efficient.

Key words: cooperation; metaphor comprehension; natural language understanding

0 引言

无论哪一种自然语言, 都普遍存在隐喻用法。隐喻是语言运用中十分普遍的现象, 不但文学语言如此, 日常语言、科技语言同样如此。隐喻思维也是人类最重要的思维机制之一, 因此这一关键问题也是人工智能技术进一步发展需要解决的中心问题之一。不解决隐喻语言的理解问题而仅仅局限于字面意义的获取, 对于语言的深入理解是远远不够的。从这个意义上讲, 找到隐喻语言理解的解决方法, 哪怕是初步的, 也必定有助于更好地提高语言的机器理解水平。

自20世纪70年代以来, 出现了一些隐喻语言理解的各种初步模型, 如基于优先语义的方法, 包括优先语义方法^[1]、词汇语义方法^[2,3]; 基于隐喻知识的方法, 包括隐喻突显理论^[4,5]、基于实例的模型^[6]和人工神经网络模型^[7]; 基于类比推理、逻辑推理的方法, 包括结构匹配^[8,9]、隐喻的结构理论(SMT)^[10]和隐喻逻辑理论^[11,12]。随着语料库语言学的发展, 利用统计方法来处理语言信息得到了越来越多的关注。除了Kintsch^[13,14]利用潜在语义分析方法从语料库中挖掘语义信息外, Mason于2004年给出了一种基于语料库的隐喻句提取引擎CorMet^[15]。

前人对隐喻做了很多研究工作, 但从合作机制出发来考察隐喻的产生与理解问题还未曾见到。本文拟从该点展开对隐喻的研究。

合作本身的一些特性与隐喻的理解有许多相似之处, 如

“同从异出”“互动”“合作双方达成一致”“达到双赢”等。本文提出基于合作机制的隐喻理解方法, 关注隐喻的理解本质, 有效地对隐喻现象进行了解读。与已有的工作相比, 本文提出的方法的主要特点有: 基于隐喻理解的互动论; 将隐喻理解建立在合作的基础上; 将基于知识的方法和基于语料库的方法有机地结合在一起; 有效地对隐喻理解内在机制进行了解读; 隐喻表述义提取的正确率较高。

1 名词性隐喻的合作机制

本文重点研究名词性隐喻的理解机制, 如“律师是狐狸”“希望的肥皂泡”“如银的月色”“祖国, 我的母亲”等。

1.1 本体和喻体的合作关系

本体和喻体的合作关系如图1所示。

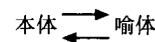


图1 本体和喻体的互动

将本体和喻体都看成是agent, 它们之间的关系也就是两个agent之间的关系。这两个agent显然是两个不同的个体, 它们之间可能合作也可能对抗。本体和喻体只有在合作的条件下才可能构成隐喻, 才具有意义。

隐喻的一个本质属性是同从异出, 即从不同的个体中看到其中的相似之处。对于本体和喻体这两个agent, 它们的合作过程就是一个求同存异的过程, 也就是同从异出。隐喻表述义的获取是隐喻研究的关键问题。可以把表述义的获取看成是

收稿日期: 2006-07-17; 修返日期: 2006-09-25 基金项目: 国家自然科学基金资助项目(60373080)

作者简介: 苏畅(1974-), 女, 博士研究生, 主要研究方向为自然语言理解和专家系统(suchang@xmu.edu.cn); 周昌乐, 教授, 博导, 主要研究方向为人工智能研究、计算语言学、理论脑科学、认知逻辑学。

合作的成果,对本体和喻体都带来好处。

1.2 合作方式

在隐喻中,本体和喻体的地位不同。本体是主动的,喻体是被动的。本体和喻体的合作方式有以下几种:

a) 喻体展示自己的属性(往往是较显著的属性),本体也具有这个属性,则一拍即合,合作成功;

b) 喻体展示自己的属性,本体不具有这个属性,但本体具有的另一个属性与喻体的属性有相似之处,则本体同意将该属性作为合作的基础;

c) 喻体展示自己的属性,本体不具有这个属性,且本体具有的所有属性与喻体的属性均无相似之处,但若本体与该属性可以搭配,则本体同意将该属性作为合作的基础。

当出现如下状况时,称本体和喻体不合作:喻体贡献出自己的属性,本体不具有这个属性,且本体具有的所有属性与喻体的属性都无相似之处,本体与该属性又不可以搭配,则合作失败。

1.3 合作所处的认知世界

在隐喻中谈合作,是针对某一认知世界的。不同的认知世界中,本体和喻体的合作可能性是不同的。在一个认知世界中,该本体与该喻体是合作的并不能保证在另一个认知世界中它们也是合作的。

2 基于合作机制的隐喻理解实验系统

2.1 前人用语料库方法实现隐喻理解系统的分析

Kintsch^[13,14]利用潜在语义分析方法从语料库中挖掘语义信息。他给出了一个基于向量空间的隐喻解释计算系统,即 CI-LSA 框架,能处理形如“X is (a) Y”的隐喻。首先,利用潜在语义分析方法(latent semantic analysis, LSA)计算语义距离,得到分别与 X 和 Y 语义相关或相似的词集合;然后建立构建集成(construction-integration, CI)模型,把与喻体 Y 语义距离相近的词语选出;与本体喻体组成反馈网络,利用上下文影响网络参数,计算出本体 X 与哪些喻体相近词具有较高的语义联结,即为隐喻的解释。Mason 给出了一种基于语料库的隐喻句提取引擎 CorMet^[15],主要关注于大规模语料库中隐喻的识别与分析。

利用潜在语义分析方法和互信息(MI)的方法先分别获取本体和喻体的特征集,再将两者取交集获取隐喻的表述义的方法在本质上是基于比较的观点,不是基于大家接受的互动论的观点。

分别获取本体和喻体的特征集,然后将两者取交集获取隐喻的表述义的方法有其固有的缺陷,原因如下:

a) 因为本体对于读者是不太熟悉的,所以在语料库中对本体的描述不多;

b) 完全采用统计方法从语料库中提取本体和喻体的属性,从现有的技术水平来看,准确率不高;

c) 本体不一定有此属性,但可以通过本体和喻体的互动,动态地生成该属性。

基于语料库的方法不受限于手工构造的知识库,所需的语

言知识是利用统计方法从大规模语料中挖掘处理的,但上述方法均未深入研究隐喻的理解机制,因此效果也不理想。本文提出方法将基于知识的方法和基于语料库的方法有机地结合在一起,创新地使用合作机制来解释隐喻行为。

2.2 系统构建的依据和原则

该系统是基于合作机制的隐喻理解实验系统。系统处理的是汉语名词性隐喻的表述义的获取问题。在喻体特征集的获取方面,发现喻体相对于本体是更为人们熟知的概念,否则无法帮助人们对本体进行认知。而本体是隐喻句的话题所在,是作者要阐述的重点,本体一般对读者来说较为陌生。

笔者手工构建了喻体特征的知识库,一方面是由于完全依据统计获得的喻体特征错误率太高,这将直接影响系统的性能;另一方面,如前所述,人们对喻体较熟悉,有能力构建这样的知识库,但构建过程中也将使用统计方法作为辅助手段。选取《读者》作为语料库,将《同义词词林》作为判别相似词的依据。

2.3 基于合作机制的隐喻表述义提取算法

2.3.1 隐喻句的形式化表示方法

隐喻句可以看成是一个三元组 $\langle T, V, F \rangle$ 。其中: T 是本体(tenor); V 是喻体(vehicle); F 为喻体的特征集, $F_v = \{c_1, c_2, \dots, c_n\}$ 是喻体的特征, $i = 1, \dots, n$ 。

2.3.2 喻体特征的同义词集

定义喻体特征 c_i 的同义词集 $M_i = \{m_j | c_j \text{ 为喻体的第 } i \text{ 个特征且 } m_j \text{ 与 } c_i \text{ 是同义词}, j = 1, \dots, t\}$

2.3.3 隐喻表述义获取算法

设本体为 x ,喻体为 y ,喻体特征集 $F_v = \{c_1, c_2, \dots, c_n\}$;对喻体的每个特征,构建对应该特征的同义词集 $M_i (i = 1, \dots, n)$ 。

算法思想:先确定选哪个特征作为待选特征,再从该特征的同义词集中选择最合适的词作为该隐喻句的表述义。

隐喻表述义获取算法如下:

a) 半手工构建喻体特征的知识库,获取喻体 y 的特征集。

b) 以《同义词词林》作为判别相似词的依据,构建特征 c_i 的同义词集 $M_i (i = 1, \dots, n)$ 。

c) 选取 M_i ,使 $F(M_i, x, y)$ 的值取最大,即 $M^* = \arg \max_{M_i} F(M_i, x, y)$ 。其中: $F(M_i, x, y)$ 由该同义词集 M_i 与本体的可搭配程度决定,使用 $F = \sum_{m_j \in M_i} P(m_j, x)$ 表示。

d) 从 M^* 中选取最合适的词作为该隐喻句的表述义。考查该同义词集中的每个特征与本体的关系,看哪个词与本体最合作。 $m_y^* = \arg \max_{m_y} P(m_y, x)$ 。

e) m_y^* 即为所求的隐喻表述义。

2.4 实验结果分析

语料库选择为《读者》,大小为 23.6 MB,体裁包括小说、散文、文摘等。之所以挑选《读者》作为语料库是基于以下原因:

a) 代表性。有一定的代表性,能反映典型的汉语语言现象。

b) 广泛性。所选语料能较全面地反映出隐喻中本体和喻体的各种用法。c) 真实性。语料自然真实。d) 比喻这种修辞方式

出现较多。

实验系统的粗框图如图 2 所示。系统采用 Visual C++ 6.0 实现,总控模块提供本系统的主界面,并可调用其他模块完成各项功能。

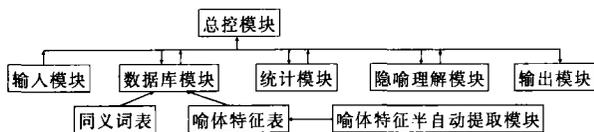


图 2 系统框图

系统处理了 100 个典型隐喻句。系统能正确理解的隐喻句举例如下:

- a) 律师是狐狸,该隐喻句表述义输出为:律师是精明的。
- b) 女人是水,该隐喻句表述义输出为:女人是温柔的。
- c) 风景如画,该隐喻句表述义输出为:风景很美。
- d) 他是大象,该隐喻句表述义输出为:他很重。
- e) 岁月如梭,该隐喻句表述义输出为:岁月流逝快。
- f) 时间是金钱,该隐喻句表述义输出为:时间宝贵。

系统无法理解个别隐喻句的原因是该语料库中有关该隐喻句的本体信息太少。隐喻句的理解正确率如表 1 所示。

表 1 隐喻句的理解正确率

隐喻句	理解正确率/%
名词性隐喻句(一阶)	92
名词性隐喻句(二阶)	80

通过实验,一方面肯定了所采用的基于合作机制的隐喻表述义提取算法的思路及其算法实现的有效性;同时也发现了个别分析不了和错误的地方,这为将来进一步完善基于合作机制的隐喻表述义提取算法指明了方向。

3 结束语

从实践中认识到,隐喻的使用是非常普遍。隐喻理解是自然语言理解中不可或缺的一环,在语篇理解中有着重要的作用。

通过观察可以发现,人们对喻体往往比对本体熟悉,这也是人们使用隐喻的原因,即通过一个较熟悉的事物去谈论另一个比较不熟悉的事物。基于此,使用半手工方式来构建喻体的属性知识库是可行的。所谓半手工方式,是指采用统计方法从语料库中提取喻体的属性,但提取后必须经过人工确认,这样才能保证属性知识库的可靠性,同时又

不乏客观性。

进一步的改进将着眼于开发基于语料库的深层语义的近义词判别系统。这能在一定程度上提高隐喻理解系统的理解正确率。

参考文献:

- [1] YORICK. A preferential pattern-seeking semantics for natural language inference[J]. *Artificial Intelligence*, 1975, 6(1):53-74.
- [2] DAN F. Met*: a method for discriminating metonymy and metaphor by computer[J]. *Computational Linguistics*, 1991, 17(1):49-90.
- [3] DAN F, YORICK W. Preference semantics, III-formedness, and metaphor [J]. *American Journal of Computational Linguistics*, 1983, 9(3-4):178-187.
- [4] WEINER E J. A Knowledge representation approach to understanding metaphor[J]. *Computational Linguistics*, 1984, 10(1):1-14.
- [5] JACOBS P S. A knowledge-based approach to language production [D]. Berkeley, CA:University of California, 1985.
- [6] MARTIN J H. A computational model of metaphor interpretation [M]. Boston:Academic Press, 1990.
- [7] VEALE T. Metaphor, memory and meaning: symbolic and connectionist issues in metaphor interpretation[D]. Dublin:Trinity College, 1995.
- [8] GENTNER D. Structure-mapping:a theoretical framework for analogy [J]. *Cognitive Science*, 1983, 7(2):155-170.
- [9] FALKENINER F. GENTENER D. The structure mapping engine:algorithm and examples[J]. *Artificial Intelligence*, 1989, 34(1):1-63.
- [10] STEINHART E C. The logic of metaphor:analogous parts of possible worlds[M]. Boston:Kluwer Academic Publishers, 2001.
- [11] 张威. 汉语语篇理解中元指代和隐喻的机器消解研究[D]. 杭州:浙江大学,2003.
- [12] 张威,周昌乐. 汉语隐喻理解的逻辑描述初探[J]. *中文信息学报*, 2004, 18(5):23-28.
- [13] KINTSCH W. Metaphor comprehension:a computational theory [J]. *Psychonomic Bulltin & Review*, 2000, 7(2):257-266.
- [14] KINTSCH W. Predication[J]. *Cognitive Science*, 2001, 25(2):173-202.
- [15] ZACHARY J M. CorMet:a computational corpus-based conventional metaphor extraction system [J]. *Computational Linguistics*, 2004, 30(1):23-44.
- [16] GOATLY A. The language of metaphors[M]. New York:Routledge, 1997.

(上接第 66 页) International Conference on the Numerical Solution of Markov Chains. 2003:267-288.

- [7] KAMVAR S, HAVELIWALA T, MANNING C. Extrapolation methods for acceleration pagerank computations[C]//Proc of International World-Wide Web Conference. 2003:221-232.
- [8] LIN S H, HO J M. Discovering informative content blocks from Web documents[C]//Proc of the 8th ACM SIGKDD International Conference. 2002:588-593.
- [9] TECAVELIWALA T H. Topic-sensitive PageRank [C]//Proc of the International World-Wide Web Conference. 2002:98-110.
- [10] WANAG Y, DeWITT D. Computing PageRank in a distributed inter-

net search system [C]//Proc of International Conference on Very Large Databases(VLDB). 2004:119-128.

- [11] LIU Bing, GROSSMAN R, ZHAI Yan-hong. Mining data records in Web pages[C]//Proc of SIGKDD. Washington DC:[s. n.], 2003:223-244.
- [12] NLPProcessor—text analysis toolkit [EB/OL]. <http://www.infogistics.com/textanalysis.html>.
- [13] ZHANG Yang, LI Zhan-bei. Improving the performance of Naïve Bayes text classifier through weakening Naïve Bayes assumption[J]. *The Computer Science*, 2002,29(8):55-63.
- [14] Open directory project[EB/OL]. <http://dmoz.org/>.